MICROBIAL SYNTHESIS OF TERPENE AND FATTY ACID BIOFUEL PRECURSORS

A Dissertation
Presented to
The Academic Faculty

by

Stephen Sarria

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Chemistry and Biochemistry

Georgia Institute of Technology
May 2018

MICROBIAL SYNTHESIS OF TERPENE AND FATTY ACID BIOFUEL PRECURSORS

Approved by:

Dr. Pamela Peralta-Yahya, Advisor
School of Chemistry and Biochemistry
Georgia Institute of Technology

Dr. Raquel Lieberman
School of Chemistry and Biochemistry
Georgia Institute of Technology

Dr. Julia Kubanek
School of Biology
Georgia Institute of Technology

Dr. Andreas Bommarius
School of Chemical and Biomolecular Engineering
Georgia Institute of Technology

Dr. Ronghu Wu
School of Chemistry and Biochemistry
Georgia Institute of Technology

Date Approved: December 13th, 2017

Para mi abuela Alba Granada

ACKNOWLEDGEMENTS

First off, I would like to thank all my family, friends, and coworkers who had a part in helping me learn and grow especially in the past five years. I am very proud of myself of having reached this moment, but I could not have done it without many great people around me.

I want to thank my parents and who have provided the best upbringing and support I could ever ask for. Both have always been supportive of my career and education and I am very grateful for all the help and guidance they have provided. I am very excited to make them proud. I especially want to thank my grandmother Alba Granada who I think about every day. Throughout my childhood she showed me the importance of hard-work and set a great example. Growing up she always said I was going to be a doctor (although she probably meant an M.D.). I am confident that she would be very happy and proud to see me finish my Ph.D.

I would also like to thank John Murnan my high school anatomy and physiology teacher who helped spark my interest in science and influenced me to pursue a college degree in Biology. I am also very grateful for the incredible professors I met while at Southern Polytechnic State University including, but not limited to, Dr. Louten, Dr. Beach, and Dr. Singh. I appreciate their help in motivating, guiding, and mentoring me towards a research-based career and into graduate school.

I especially want to thank my thesis advisor Pamela Peralta-Yahya. Since the beginning of graduate school, she has always been very helpful and has provided great

career and life advice. She has been essential in molding me into a better critical thinker and overall scientist while also keeping me motivated during difficult times I encountered in research. It has been a great experience being a part of the lab set up in 2012 and helping it progress through the years.

Lastly, I want to express my appreciation to all the great lab mates I met and worked with throughout my time at Georgia Tech. Amy Ehrenworth was a great lab mate to work for five years. We were the first graduate students to set up the Peralta-Yahya lab and experienced the crazy ups and downs of graduate school together. I also appreciate working with a former post-doc Kuntal Mukherjee who taught me a lot about yeast and Emily Yasi who was great to work with and was always there for help. Lastly, I would like to thank Adam Verga a very talented undergraduate student who helped generate data needed for two manuscripts.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS iv

LIST OF TABLES ix

LIST OF FIGURES x

LIST OF SYMBOLS AND ABBREVIATIONS xvi

SUMMARY xvii

CHAPTER 1. Microbial synthesis of pinene 1
1.1 Abstract 1
1.2 Introduction 1
1.3 Results 6
1.3.1 Identifying pinene synthases 6
1.3.2 Microbial production of pinene (co-expression) 7
1.3.3 Geraniol toxicity 12
1.3.4 Pinene toxicity 14
1.3.5 Overcoming GPPS inhibition by GPP via protein fusions 14
1.3.6 Pinene microbial production using protein fusions 16
1.3.7 Rationalization of operon vs. protein fusion performance 17
1.3.8 Ratios of α- and β-pinene 19
1.4 Discussion 22
1.5 Methods 24
1.5.1 Plasmid construction 25
1.5.2 Pinene production and quantification 28
1.5.3 Pinene and geraniol toxicity measurements 29
1.5.4 Pinene conversion using cell lysate 29
1.5.5 Flux balance analysis yield calculations 30
1.5.6 Protein quantification protocol 31
1.6 References 32

CHAPTER 2. Microbial synthesis of medium-chain chemicals from renewables 37
2.1 Abstract 37
2.2 Introduction 37
2.3 Medium-chains in nature 43
2.4 Engineering medium-chain chemical precursors 44
2.4.1 Fatty acid biosynthesis 45
2.4.2 Reversed β-oxidation pathway 46
2.4.3 Terpene biosynthesis 47
2.4.4 Amino acid biosynthesis 47
2.5 From medium-chain precursors to products 47
2.5.1 Fatty acids 48
2.5.2 Alkanes 50

2.5.3 α-olefins 51
2.5.4 Alkyl esters 51
2.5.5 ω-hydroxyacids 53
2.5.6 α, ω-diacids 54
2.5.7 Alcohols 54
2.5.8 Methyl ketones 55
2.6 Outlook 56
2.7 References 60

CHAPTER 3. Matching protein interfaces for improved medium-chain fatty acid production 67
3.1 Abstract 67
3.2 Introduction 68
3.3 Results 71
3.3.1 Screening medium-chain acyl-ACP TEs and *E. coli* hosts for MCFA production 71
3.3.2 Engineering AbTE for improved MCFA titers 73
3.3.3 Extending cultivation time to increase MCFA titers 80
3.3.4 Detecting MCFAs secreted by engineered *E. coli* using a GPCR-based MCFA sensor 81
3.4 Discussion 83
3.5 Materials and Methods 85
3.5.1 Reagents 85
3.5.2 Plasmid construction 85
3.5.3 AbTE mutant generation 86
3.5.4 Fatty acid production and quantification 86
3.5.5 E. coli secreted MCFA detection by *S. cerevisiae* MCFA sensor 87
3.5.6 Homology model and docking 88
3.5.7 SDS-PAGE gel 88
3.6 References 89

CHAPTER 4. Phylogeny-based engineering of a bacterial acyl-ACP thioesterase 92
4.1 Abstract 92
4.2 Introduction 92
4.3 Results 95
4.3.1 Multiple sequence alignment of *A. baylyi* sequence 96
4.3.2 *In vivo* activity screen of phylogeny-based AbTE mutants 98
4.4 Discussion 102
4.5 Materials and Methods 104
4.5.1 Reagents 104
4.5.2 Multiple sequence alignment and variability analysis 105
4.5.3 AbTE mutant generation 105
4.5.4 Fatty acid production and quantification 106
4.5.5 *A. baylyi* thioesterase homology model and docking 106
4.5.6 SDS-PAGE gel 107
4.6 References 107

CHAPTER 5. Future outlook **110**
5.1 Conclusions and Future outlook **110**
5.1.1 Improving microbial titers of pinene to be cost-competitive 110
5.1.2 High-throughput screening of medium-chain fatty acid producer strains 111
5.2 References **112**

APPENDIX A. Plasmid tables **113**
Table A-1: Chapter 1 plasmids **113**
Table A-2: Chapter 3 plasmids **117**
Table A-3: Chapter 4 plasmids **119**

APPENDIX B. Strain tables **122**
Table B-1: Chapter 1 strains **122**
Table B-2: Chapter 3 strains **122**
Table B-3: Chapter 4 strains **124**

APPENDIX C. Primer tables **126**
Table C-1: Chapter 1 primers **126**
Table C-2: Chapter 3 primers **128**
Table C-3: Chapter 4 primers **129**

APPENDIX D. Construction of Site directed mutants **134**
Table D-1: Chapter 3 site directed mutagenesis templates and primers **134**
Table D-2: Chapter 4 site directed mutagenesis templates and primers **134**

APPENDIX E. theoretical yield calculations **137**
E.1 Calculations of stoichiometric theoretical yields of medium-chain chemicals **137**
E.2 Calculations of theoretical yield of medium-chain chemicals **139**
E.3 References **148**

APPENDIX F. DNA sequences **150**
F.1 Chapter 1 sequences **150**
F.1.1 *Abies grandis* geraniol diphosphate synthase 150
F.1.2 *Picea abies* geraniol diphosphate synthase 151
F.1.3 *Pinus taeda* geraniol diphosphate synthase 151
F.1.4 *Abies grandis* pinene synthase 152
F.1.5 *Picea abies* pinene synthase 154
F.1.6 *Pinus taeda* pinene synthase 155
F.2 Chapter 3 sequences **157**
F.2.1 *Acinetobacter baylyi* thioesterase 157
F.2.2 *Cocos nucifera* thioesterase FatB3 158
F.2.3 *Cuphea palustris* thioesterase FatB1 159
F.2.4 *Umbellularia californica* thioesterase FatB2 160

LIST OF TABLES

Table 2-1	Medium-chain chemical uses and highest titers. Prices were obtained from Sigma-Aldrich.	38
Table 2-2	Enzymes engineered for MC chemical production. a. *In vitro*	42
Table 2-3	Theoretical Yield (TY), highest reported yields and titers of microbially produced MC chemicals. a. Molar yield from glucose calculated using COBRA toolbox for Matlab. b. Calculated as (moles of product)/(moles of carbon source) c. Theoretical stoichiometric yield is based on carbon chain length of MCFA precursor, does not use carbon from glucose for ethyl side chain. d. Theoretical pathway, involves undiscovered enzymes.	57
Table 3-1	Saturated fatty acid percent composition produced by AbTE:WT and variants.	79
Table 4-1	*A. baylyi* thioesterase amino acids selected for mutagenesis	97
Table 4-2	Chain-length profile of best MCFA-producing *A. baylyi* TE phylogeny-based mutants.	101

LIST OF FIGURES

Figure 1.1 Energy density of petroleum-based fuels and advanced biofuels. Shown is the heating value of petroleum-based fuels (black) and advanced biofuels (green) as a function of density. Current advanced biofuels have lower density and heating value when compared to high-energy density petroleum-based fuels such as JP-10 and RJ-5. Pinene dimers (red) have similar density and heating value to JP-10. Pinene dimers mimic the strained ring systems found in JP-10 and RJ-5. Pinene dimers can be generated via pinene (red) dimerization using chemical catalysis. Data to generate this graph were obtained from the Biomass Energy Data Book 2011 and ref 11. 3

Figure 1.2 Microbial production of pinene, the immediate precursor to a high-energy density biosynthetic tactical fuel. *a*. *Escherichia coli* (yellow box) converts simple sugars into acetyl-CoA via primary metabolism. Introduction of a heterologous mevalonate pathway converts acetyl-CoA into isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP). Addition of geranyl diphosphate synthase (GPPS) converts IPP and DMAPP into geranyl diphosphate (GPP), which is cyclized by pinene synthase (PS) to produce pinene. Pinene can be dimerized into pinene dimers via chemical catalysis to generate a high energy density biosynthetic tactical fuel. b. Construct design. The mevalonate pathway is present in one plasmid as an operon under control of the LacUV5 promoter. The PS and GPPS are present in separate plasmids as an operon under control of the Trc promoter. Also shown is the feedback regulation at the end of the pinene biosynthetic pathway. GPP inhibits GPPS and PS, while pinene inhibits PS. c. Pinene synthase cyclization mechanism of GPP to pinene. 6

Figure 1.3 Microbial production of pinene via co-expression of geranyl diphosphate synthase and pinene synthase. a. Pinene titers as a function of the GPPS and PS co-expressed. b. Toxicity of geraniol, a proxy for GPP, and farnesol, a proxy for FPP, on cell growth. Cell growth was normalized to the growth of the same strain in the absence of geraniol or farnesol. c. Toxicity of intracellularly-produced GPP. Cell growth was normalized to the growth of the same strain not expressing GPPS. The experiments shown in a., b., and c. were done in triplicate and the error bars represent the standard deviation from the mean. 8

Figure 1.4 Pinene specific production using protein co-expression and protein fusions. The cultures were grown at 37°C until they reached OD_{600}=0.8 before induction with 1mM IPTG and overlayed with 20% dodecane. Pinene titers and cell growth were measured after 72hrs. Shown is the average of three experiments, the error bars represent the standard deviation from the mean. 9

Figure 1.5 Cell growth after 72hrs of pinene microbial production. The cultures were grown at 37°C for 3 hrs before induction with 1mM IPTG and overlayed with 20% dodecane. Cell growth was measured after 72hrs. Shown is the average of three experiments and the error bars represent the standard deviation from the mean. 10

Figure 1.6 Specific growth rate of the strains used in the pinene microbial production. The cultures were grown in EZ-rich media with 1% glucose at 37°C for 3 hrs before induction with 1mM IPTG, and grown until they reached saturation (20hrs). The control is *E. coli* MG1655 harboring both an ampicillin and a chloramphenicol vector. 10

Figure 1.7 SDS-PAGE gel and western blot of pinene synthases (PSs). SDS-PAGE (4-12% BisTris) and western blot were run simultaneously with equal loading. Lanes: 1. *Picea abies* PS (65 kD) insoluble fraction, 2. *Abies grandis* PS (67 kD) insoluble fraction, 3. *Pinus taeda* PS (70 kD) insoluble fraction, 4. Ladder, 5. *Picea abies* PS soluble fraction, 6. *Abies grandis* PS soluble fraction, 7. *Pinus taeda* PS soluble fraction. Western Blot has the same line arrangement as coomassie stained SDS-PAGE gel. 11

Figure 1.8 SDS-PAGE gel and western blot of geranyl diphosphate synthases (GPPSs). SDS-PAGE (4-12% BisTris) and western blot were run simultaneously with equal loading. Lanes: 1. *Picea abies* GPPS (32 kD) insoluble fraction, 2. *Abies grandis* GPPS (32 kD) insoluble fraction, 3. *Pinus taeda* GPPS (32 kD) insoluble fraction, 4. Ladder, 5. *Picea abies* GPPS soluble fraction, 6. *Abies grandis* GPPS soluble fraction, 7. *Pinus taeda* GPPS soluble fraction. Western Blot has the same line arrangement as coommasie stained SDS-PAGE gel. 12

Figure 1.9 Microbial production of pinene via fusion of geranyl diphosphate synthase and pinene synthase. a. Fusion protein design. GPPS at the N-terminus of the fusion results in GPPS and PS active sites facing one another. PS at the N-terminus of the fusion results in GPPS and PS facing in the same direction. Peach: *M. spicata* limonene synthase structure (PDB: 2ONG). Magenta: *M. pipetira* GPPS (PDB: 3OAC). Green and cyan: PS fused to GPPS. b. Pinene titers of protein fusions with three, six, and nine amino acid 16

linker lengths. c. Microbial pinene titers as a function of the GPPS and PS protein fusions. The experiments shown in b. and c. were done in triplicate, and the error bars represent the standard deviation from the mean.

Figure 1.10 Analysis of GPP inhibition of pinene synthase alone and when part of a protein fusion with geranyl diphosphate synthase. A. Percent conversion of GPP to pinene in cell extract of *E. coli* expressing the *A. grandis* PS (PS) alone or the *A. grandis* geranyl diphosphate synthase-pinene synthase (AgGPPS-(GSG)2-AgPS) protein fusion in the presence of manganese. B. Percent conversion of GPP to pinene in cell extract of *E. coli* expressing the *A. grandis* PS (PS) alone or the *A grandis* geranyl diphosphate synthase-pinene synthase (AgGPPS-(GSG)2-AgPS) protein fusion in the presence of magnesium. Percent conversion was calculated by dividing the moles of pinene produced over the moles of GPP at the beginning of the reaction. The experiments were done in triplicate, and the error bars represent the standard deviation from the mean. 19

Figure 1.11 Microbially-produced pinene isomer ratios as a function of geranyl diphosphate synthase (GPPS) and pinene synthase (PS). a. Ratio of α:β-pinene when matching different GPPSs and PSs using either operons of fusions. b. Ratio of α:β-pinene in cell extract of *E. coli* expressing *A. grandis* PS with either manganese or magnesium as the cofactor. c. Homology model of A. grandis GPPS fused to *A. grandis* PS. The N-terminal strand and the F-K helix and loop of GPPS are highlighted in red. The experiments shown in a. and b. were done in triplicate, and the error bars represent the standard deviation from the mean. 22

Figure 2.1 Production of chemicals from lignocellulosic biomass. Lignocellulosic biomass comprises the glucose polymer cellulose, hemicellulose and lignin. Depending on the biomass source, cellulose can constitute from 90% (cotton) to 50% (wood) of the biomass. The ease of depolymerization of cellulose by cellulase into glucose monomers, and the fact that it forms the largest portion of lignocellulosic biomass, makes it the main substrate for the microbial production of chemicals. Hemicellulose, a polymer composed of C6 and C5 sugars, is challenging to depolymerize due to the different reactivity of the linkages in the polymer. The main component of hemicellulose is the C5 sugar, which can be naturally used for fermentation and chemical production by Escherichia coli[10], but not *Saccharomyces cerevisiae*. About 35% of lignocellulosic biomass is composed of lignin, a heterogenous phenolic polymer, which is increasingly being used to produce chemicals. Lignin monomers, such as coumaric acid and ferulic acid, can be funneled to key metabolic intermediates, such as 41

acetyl-CoA, to produce value added chemicals in *Pseudomonas*[11]. Key challenges in the deconstruction of lignocellulosic biomass to simple sugars include the efficient and cost-effective delignification of biomass, the cost and activity of enzymes to break down cellulose and hemicellulose into sugar monomers, and the discovery and engineering of enzymes able to break down lignin into lignin monomers. Plant biomass is comprised of varying proportions of lignin, glucose and cellulose.

Figure 2.2 Precursor pathways for MC chemicals. Fatty acid biosynthesis (blue) and modified or reversed β-oxidation pathway (orange) produce C8-C12 acyl-CoA/ACP intermediates (n= 4, 6, 8). Precursors are extended two carbon units at a time through condensation of acetyl-CoA (reversed β-oxidation) or malonyl-CoA (fatty acid biosynthesis) followed by reduction, dehydration, and a second reduction. Boxed intermediates are key precursors to MC chemicals. 45

Figure 2.3 Metabolic pathways for medium-chain (MC) chemicals. Fatty acid biosynthesis (blue), and modified and reversed β-oxidation pathways (orange) generate fatty acyl-intermediate precursors that are modified by chain-tailoring enzymes (green) into MC chemicals. Other MC chemical producing pathways are amino acid and terpene biosynthesis (purple). n= 5-9 for all MC chemicals except for alkanes, α-olefins, and α,ω-diacids where n = 4-8. 48

Figure 3.1 Medium-chain fatty acid biosynthesis in *Escherichia coli*. A. *E. coli* Type II fatty acid synthase (FAS) extends and reduces an acyl chain bound to acyl-carrier protein (ACP). All enzymes in FAS interact with ACP. Thioesterases (TEs) hydrolyze acyl-ACPs to free fatty acids of different chain lengths according to their substrate specificity. B. Matching the surface interface between the native *E. coli* ACP and medium-chain heterologous thioesterases, *such* as *Acinetobacter baylyi* thioesterase, improves medium-chain fatty acid production in *E. coli*. 71

Figure 3.2 Medium-chain fatty acid production and calculated *E. coli* ACP interphase with thioesterases. A. Secreted fatty acid titers of *E. coli* (MG1655) expressing four heterologous thioesterases. B. Secreted fatty acid titers by different *E. coli* strains expressing *Acinetobacter baylyi* thioesterase (AbTE wt). C. Secreted and total (secreted plus intracellular and membrane bound) fatty acid titers produced by *E. coli* expressing AbTE:WT and inactive AbTE (AbTE:S11A). The experiments were done in triplicate and the error bars represent the standard deviation from the mean. D. Docking of *E. coli* ACP (magenta, PDB ID: 1FAE*)* and *E. coli* 'TesA (cyan, PDB ID 1U8U) identifies potential residues on the 'TesA surface that are important 72

for interactions between ‘TesA and ACP. E. Homology model of AbTE with surface residues equivalent to ‘TesA labelled.

Figure 3.3 Percent sequence identity between thioesterases used in this work. Bacterial thioesterases (blue). Plant thioesterases (green). 73

Figure 3.4 Gas chromatograms of AbTE and variants. A. AbTE:S11A (inactive enzyme), B. Wild-type AbTE) C. AbTE:G17R D. AbTE:G17R/A165R. Single Ion Monitoring: 74 and 87. 74

Figure 3.5 Amino acid sequence alignment of AbTE and *E. coli* ‘tesA. Red boxes indicate amino acids targeted for mutagenesis. 75

Figure 3.6 Secreted fatty acid and protein levels of *E. coli* expressing *Acinetobacter baylyi* thioesterase (AbTE) and AbTE mutants. A. Secreted fatty acid levels of *E. coli* expressing AbTE single, double, and triple arginine mutants, as well as AbTE single and double glutamate mutants. All experiments were done in triplicate and the error bars represent the standard deviation from the mean. B. Left: Coomassie stained SDS-PAGE gel of induced (+IPTG) and uninduced (-IPTG) *E. coli* cultures expressing AbTE:WT, AbTE:G17R and AbTE:G17R/A165R. Right: Serial dilution of induced *E. coli* cultures expressing AbTE:WT, AbTE:G17R and AbTE:G17R/A165R. 77

Figure 3.7 Remaining AbTE double mutant combinations. Secreted fatty acid titers produced by E. coli expressing AbTE, AbTE:T120R/A121R, AbTE:T120R/A165R, and AbTE:A121R/A165R. All experiments were done in triplicate and the error bars represent the standard deviation from the mean. 78

Figure 3.8 Cultivation time effect on medium-chain fatty (MCFA) acid titers and MCFA detection in the producer supernatant via a GPCR-based MCFA sensor. A. Time course of saturated fatty acid titers produced by *E. coli* expressing AbTE, AbTE:G17R, and AbTE:G17R/A165R. B. Schematic of *E. coli* produced MCFA by a GPCR-based MCFA sensor. MCFA producer: glucose is fed to *E. coli* expressing a AbTE to secrete MCFAs to the supernatant. MCFA sensor: *S. cerevisiae* expressing the GPCR OR1G1 (blue) detects MCFAs in the producer supernatant resulting in expression of green fluorescent protein. D. Detection of MCFAs by the MCFA sensor in the *E. coli* supernatant. Sensor GFP fluorescence and MCFA (C8-C12) titers as a function of AbTE variant expressed in E. coli. 81

Figure 3.9 MCFA *E. coli* producer cell and *S. cerevisiae* sensor pair. MCFAs produced by the producer are detected by the sensor via a 82

fluorescent readout.

Figure 3.10 Effect of fusing maltose binding protein (MBP) to AbTE mutants. Gas chromatograms of MBP-AbTE:G17R (left) and MBP-AbTE:G17R/A165R (right). Single Ion Monitoring: 74 and 87. 83

Figure 4.1 AbTE multiple sequence alignment, Logo representation. Positions with only one letter are conserved while those with multiple letters are variable. Letter size corresponds to the most dominant residue in that position. Positions targeted for mutation contain an asterisk. Sequence logo created using http://weblogo.berkeley.edu/logo.cgi. 96

Figure 4.2 Phylogeny-based engineering of *Acinetobacter baylyi* thioesterase (AbTE). A. AbTE homology model with docked C12 acyl-phosphopantetheine (green) and variable residues targeted for mutagenesis. B. Secreted fatty acid titers of *E. coli* expressing AbTE single mutants. Green represents medium-chain fatty acids (MCFAs, C8-C12) and purple represents long-chain fatty acids (LCFAs, C14-C18). Asterisk corresponds to mutants with MCFA titer titers with statistically significant improvement, *P*-value < 0.05 (two-tailed t test). All experiments were done in triplicate and the error bars represent the standard deviation from the mean. 99

Figure 4.3 Coomassie stained SDS-PAGE gel of induced *E. coli* cultures expressing AbTE:WT and best variants from single mutant screen AbTE:S91K, AbTE:E94S, AbTE:T120A, and AbTE:S154D. 100

Figure 4.4 Combination of phylogeny-based *A. baylyi* TE mutants. A. AbTE homology model with docked C12 acyl-phosphopantetheine (green). Residues shown to improve titers from single mutant screen are shown in yellow and previously mutated interface residues (white). B. Secreted fatty acid levels of *E. coli* expressing best AbTE single mutants and mutant combinations. Green represents medium-chain fatty acids (MCFAs, C8-C12) and purple represents long-chain fatty acids (LCFAs, C14-C18). Asterisk corresponds to mutant with MCFA titers with statistically significant improvement, *P*-value < 0.05 (two-tailed t test). All experiments were done in triplicate and the error bars represent the standard deviation from the mean. 101

Figure 4.5 Secreted total fatty acid (C8-C18) titers of *E. coli* expressing AbTE single mutants. Asterisk corresponds to mutant with fatty acid titers with statistically significant improvement, *P*-value < 0.05 (two-tailed t test). All experiments were done in triplicate and the error bars represent the standard deviation from the mean. 103

LIST OF SYMBOLS AND ABBREVIATIONS

IPP	Isopentenyl diphosphate
DMAPP	Dimethyl allyl diphosphate
GPP	Geraniol diphosphate
PS	Pinene synthase
GPPS	Geraniol diphosphate synthase
GC-MS	Gas chromatography-mass spectrometry
MC	Medium-chain
FAS	Fatty acid synthase
KS	Ketosynthase
MPT	Malonyl-palmitoyl transferase
AT	Acetyl transferase
ACP	Acyl-carrier protein
ADO	Aldehyde deformylating oxygenase
WS	Wax-ester synthase
OMT	*O*-methyl transferase
ADH	Alcohol dehydrogenase
ALDH	Aldehyde dehydrogenase
MCFA	Medium-chain fatty acids
TE	Thioesterase
AbTE	*A. baylyi* thioesterase

SUMMARY

Our dependence on petroleum-derived chemicals and fuels has accelerated the need for alternative platforms for renewable, sustainable, and greener chemical synthesis. The engineering of microorganisms is a promising alternative to producing chemicals from renewables that are traditionally obtained from petroleum. In recent years, much progress has been achieved in the development of technologies to aid the engineering of microbes to reach industrially relevant chemical titers. Here, I provide a body of work that encompasses the microbial production of terpene and fatty acid -based chemicals that have applications as precursors to fuels and commodity chemicals.

Chapter 1 involves the microbial synthesis of pinene, a plant-derived terpene natural product. Pinene can be chemically converted into a pinene dimer that contains similar energy density to JP-10, a military fuel used for aircraft launched missiles. To provide a sustainable source of pinene, we engineered *Escherichia coli* for pinene production. We combinatorially expressed three pinene synthases (PS) and three geranyl diphosphate synthases (GPPS), with the best combination achieving ~28 mg/L of pinene. We speculated that pinene toxicity was limiting production; however, toxicity should not be limiting at current titers. Because GPPS is inhibited by geranyl diphosphate (GPP) and to increase flux through the pathway, we combinatorially constructed GPPS-PS protein fusions. The *Abies grandis* GPPS-PS fusion produced 32 mg/L of pinene, a 6-fold improvement over the highest titer previously reported in engineered *E. coli*. Finally, we investigated the pinene isomer ratio of our pinene-producing microbe and discovered that the isomer profile is determined not only by the identity of the PS used but also by the

identity of the GPPS with which the PS is paired. We demonstrated that the GPP concentration available to PS for cyclization alters the pinene isomer ratio.

Chapter 2 is a review of the microbial synthesis of medium-chain chemicals from renewables. Chemicals containing a linear, medium-chain (C8-C12) hydrocarbon are used as fuels, commodity and specialty chemicals. Fatty acid synthesis and reversed β-oxidation are the major source of to medium-chain chemical precursors, but chain lengths vary, such that desired products must be purified, and titers rarely exceed milligrams of product per liters. As medium-chain chemicals are not endogenous to industrially relevant microbes, researchers go beyond the overexpression of endogenous enzymes and deletion of competing pathways to produce these chemicals. Scientists are engineering the substrate specificity and activity of pathway enzymes that generate the fatty acyl intermediates and chain tailoring enzymes to boost the scope of products, and the yields, of medium-chain chemicals. A gap in knowledge exists in chain tailoring enzymes structures and kinetics, as well as rapid assays to screen for medium-chain chemical levels. Filling this gap and developing appropriate technologies will enable us in to achieve g/L levels of these chemicals.

In **Chapter 3** we engineered the heterologous medium-chain thioesterase from *Acinetobacter baylyi* (AbTE) for the improved production of medium-chain fatty acids (MCFAs) by matching the interphase of the heterologous thioesterase and AbTE. Medium-chain fatty acids (MCFAs) are key intermediates in the synthesis of medium-chain chemicals including α-olefins and dicarboxylic acids. In bacteria, microbial production of MCFA is limited by the activity and product profile of acyl-ACP thioesterases. Here, we engineer a heterologous bacterial acyl-ACP thioesterase for

improved MCFA production in *Escherichia coli*. Electrostatically matching the interface between the heterologous medium-chain *Acinetobacter baylyi* acyl-ACP thioesterase (AbTE) and the endogenous *E. coli* ACP by replacing small nonpolar amino acids on the AbTE surface for positively charged ones increased secreted MCFA titers more than three-fold. Replacing the same small nonpolar amino acids with negatively charged ones reduces MCFA titers by half. After extending the cultivation time to 72 hours, the best AbTE mutant produced 131 mg/L of MCFAs, and MCFAs are 80% of all secreted fatty acid chain lengths. Finally, a previously developed G-protein coupled receptor (GPCR)-based MCFA sensor was used to detect MCFA levels secreted by *E. coli* expressing the AbTE variants. This work demonstrates that engineering the interface between heterologous proteins and interacting host proteins is a promising approach to increase the titers of microbially-produced chemicals. Further, this work validates the GPCR-based MCFA sensor for detection of chemicals in the producer microbe's supernatant setting the stage for the sensor-guided engineering of MCFA producing microbes.

In **Chapter 4**, we targeted the engineering of the *A. baylyi* thioesterase using an alternative, phylogeny-based approach. Previous work suggests sequence variability in acyl-ACP thioesterases corresponds to the differences in specificity observed in the phylogeny. We hypothesized that mutations of variable positions on AbTE to those residues found in the phylogeny could improve AbTE specificity/activity towards MCFAs. Alignment of AbTE related sequences from the *Acinetobacter* genus identified variable positions at the thioesterase surface. The phylogeny-based single mutant screen revealed two surface residue mutations that improved MCFA titers ~1.3-fold over wild-type. Combination of the phylogeny-derived AbTE mutations did not improve MCFA

titers over the single mutants, but led to a shift in fatty acid profile toward long-chain fatty acids. The phylogeny-based mutagenesis was marginally successful in improving MCFA titers.

Chapter 5 summarizes the conclusions obtained by this work while exploring the future outlook of each of the previous chapters.

CHAPTER 1. MICROBIAL SYNTHESIS OF PINENE

1.1 Abstract

The volumetric heating values of today's biofuels are too low to power energy-intensive aircraft, rockets, and missiles. Recently, pinene dimers were shown to have a volumetric heating value similar to that of the tactical fuel JP-10. To provide a sustainable source of pinene, we engineered *Escherichia coli* for pinene production. We combinatorially expressed three pinene synthases (PS) and three geranyl diphosphate synthases (GPPS), with the best combination achieving ~28 mg/L of pinene. We speculated that pinene toxicity was limiting production; however, toxicity should not be limiting at current titers. Because GPPS is inhibited by geranyl diphosphate (GPP), and to increase flux through the pathway, we combinatorially constructed GPPS-PS protein fusions. The *Abies grandis* GPPS-PS fusion produced 32 mg/L of pinene, a 6-fold improvement over the highest titer previously reported in engineered *E. coli*. Finally, we investigated the pinene isomer ratio of our pinene-producing microbe and discovered that the isomer profile is determined not only by the identity of the PS used, but also by the identity of the GPPS with which the PS is paired. We demonstrated that the GPP concentration available to PS for cyclization alters the pinene isomer ratio.

1.2 Introduction

Advanced biofuels have properties similar to petroleum-based fuels and can be "dropped in" to the existing transportation infrastructure[1]. Recent progress in engineering microbes for the production of advanced biofuels has resulted in biosynthetic alternatives to gasoline, such as butanol[2]; diesel, such as fatty acid ethyl esters[3]; and diesel precursors, such as bisabolene[4] and farnesene[5]. However, the development of microbial platforms for the production of high-energy density fuels, i.e., tactical fuels for use in aircraft and aircraft-launched missiles, has lagged behind. Current biosynthetic jet fuels, such as hydroprocessed esters and fatty acids (HEFA), derived from the natural oils present in oil-seed plants such as Camelina[6] and algae triglyceride[7], have been used to power both military and commercial aircraft in 50:50 blends with Jet-A fuel[8]. More recently, dehydration of butanol into butene followed by oligomerization resulted in butene oligomers that can also be used as jet fuel[9-10]. Existing biosynthetic jet fuels, however, lack the volumetric energy content required to replace high-energy density fuels such as the tactical fuels JP-10, tetrahydrodicyclopentadiene, and RJ-5, a mixture of norbornadiene dimers used for aircraft-launched missiles (Figure 1.1). Attaining the volumetric energy content necessary for tactical fuels requires mimicking the strained ring systems found in JP-10 and RJ-5. Recently, pinene dimers have been shown to contain high volumetric energy similar to that found in JP-10[11]. Pinene dimers are synthesized via chemical dimerization of pinene, a bicyclic terpene[11].

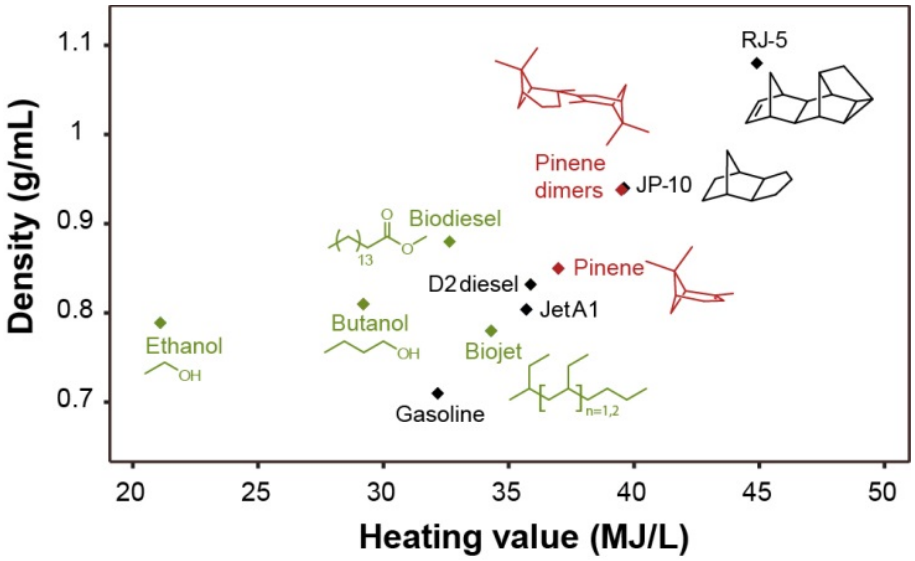

Figure 1.1: Energy density of petroleum-based fuels and advanced biofuels. Shown is the heating value of petroleum-based fuels (black) and advanced biofuels (green) as a function of density. Current advanced biofuels have lower density and heating value when compared to high-energy density petroleum-based fuels such as JP-10 and RJ-5. Pinene dimers (red) have similar density and heating value to JP-10. Pinene dimers mimic the strained ring systems found in JP-10 and RJ-5. Pinene dimers can be generated via pinene (red) dimerization using chemical catalysis. Data to generate this graph were obtained from the Biomass Energy Data Book 2011 and ref 11.

Terpenes, such as pinene, are plant natural products with a wide range of functions from defense to pollinator attractants and allelopathy compounds[12]. More than 1,000 naturally occurring monoterpenes have been identified[13], and monoterpenes play commercial roles as flavors, fragrances, insecticides, and pharmaceuticals[14]. Today, the major source of pinene is turpentine, a byproduct of the wood pulp industry[15]. In plants, monoterpenes (C10) such as pinene are biosynthesized in the plastid from the C5 intermediates isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP) generated via the deoxyxylulose-5-phosphate (DXP) pathway. Geranyl diphosphate synthase (GPPS) carries out a head-to-tail condensation of IPP and DMAPP to produce geranyl diphosphate (GPP, C10), which is, in turn, cyclized by pinene synthase (PS) to

produce either α- or β-pinene. In contrast, sesquiterpenes (C15), such as bisabolene, are produced via the mevalonate pathway in the cytosol. In the cytosol, IPP and DMAPP are condensed by farnesyl diphosphate synthase (FPPS) to produce farnesyl diphosphate (FPP, C15), which is, in turn, cyclized by sesquiterpene synthases into a variety of sesquiterpenes.

Given the large quantities of pinene dimers needed for use as a biofuel, engineering microorganisms to produce pinene from inexpensive sugars may be the most convenient and cost-effective approach to obtaining the necessary quantities of this advanced biofuel precursor. Using an engineered *Escherichia coli* strain for the overproduction of IPP and DMAPP via the mevalonate pathway[16], we previously demonstrated pinene production of ~1 mg/L from ionic liquid-treated switchgrass after introduction of a GPPS and PS[17]. More recently, *E. coli* was engineered to produce pinene at titers of ~5 mg/L in shake flasks using a complex medium (beef broth) containing 2% glucose[18], achieved in one of seven tested beef broths, the remainder of which resulted in ~1 mg/L. Notably, microbial pinene titers are orders of magnitude lower than those of sesquiterpenes (bisabolene[4]) and diterpenes (taxadiene[19]). Further, a different monoterpene, limonene, has been produced microbially at a titer of ~400 mg/L[20]. Likely reasons for the low microbial production of pinene may be 1) the toxicity of pinene or GPP to *E. coli*[21]; 2) the inhibition of GPPS by its substrate (GPP)[22] or high concentration of magnesium[23-24]; and 3) inhibition of PS by its substrate (GPP)[25] and product (pinene)[26], or reduced activity when magnesium is used as a cofactor for catalysis rather than manganese[26].

Here, we engineered *E. coli* for the production of pinene, the immediate precursor to pinene dimers (Figure 1.2). Our strategy relied on combinatorially screening for high-

flux PS and GPPS enzymes for the last two steps of the pathway. This was followed by protein fusion of GPPS to PS to reduce GPP inhibition of GPPS activity and to relieve potential GPP toxicity to the *E. coli* host. Using a previously engineered *E. coli* for the overproduction of IPP and DMAPP[17], we combinatorially screened high-fidelity GPPS and PS enzymes from different plant conifers, resulting in strains that produced between 11.2 mg/L and 27.9 mg/L of pinene. To reduce GPP inhibition of GPPS activity, we constructed GPPS-PS protein fusions combinatorially, yielding pinene titers between 11.4 mg/L and 32.4 mg/L. Given the limited improvement obtained via GPPS-PS enzyme fusions, we speculated that the toxicity of pinene to the host was limiting production. We discovered that α-pinene is more toxic than β-pinene; however, pinene toxicity should not be limiting growth and pinene production at current titers. Finally, we investigated the pinene isomer ratio produced by our engineered *E. coli,* given the industrial desirability of β-pinene. We discovered that the pinene isomer profile is determined not only by the identity of the PS, but also by the identity of the GPPS with which the PS is paired, and whether the enzymes are co-expressed or fused. We also demonstrated that the GPP concentration available to PS for cyclization alters the pinene isomer ratio.

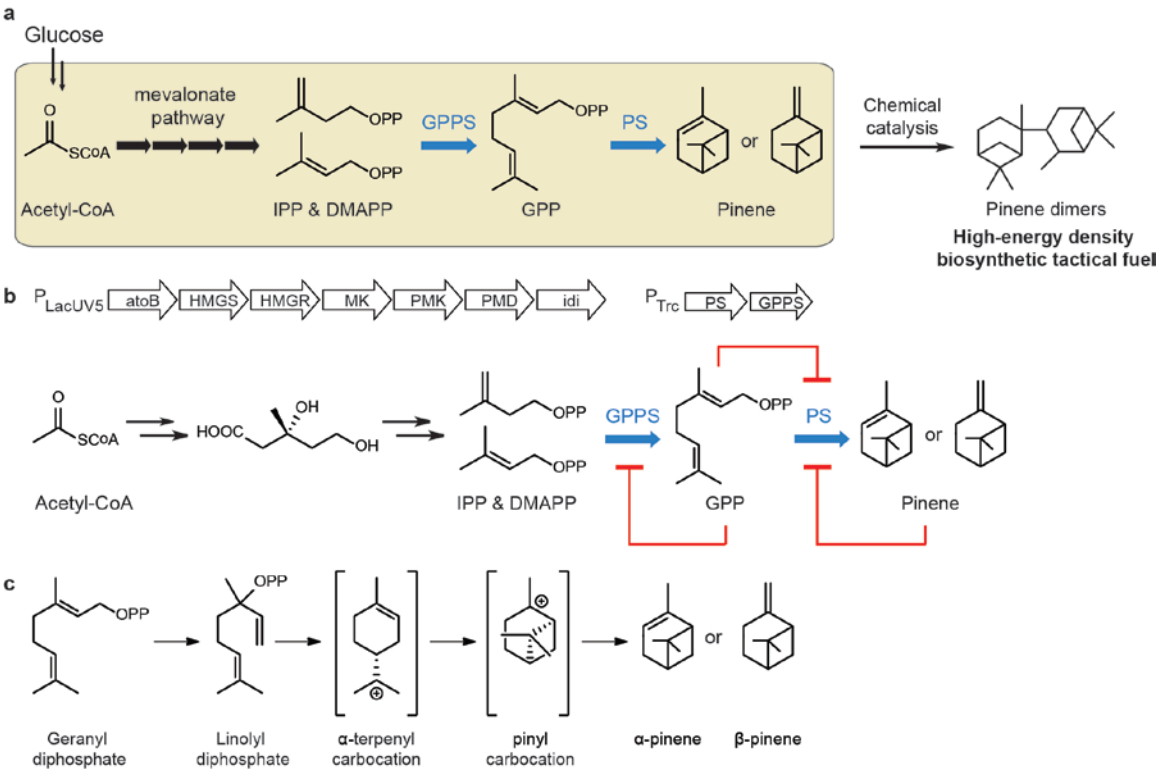

Figure 1.2: Microbial production of pinene, the immediate precursor to a high-energy density biosynthetic tactical fuel. a. *Escherichia coli* (yellow box) converts simple sugars into acetyl-CoA *via* primary metabolism. Introduction of a heterologous mevalonate pathway converts acetyl-CoA into isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP). Addition of geranyl diphosphate synthase (GPPS) converts IPP and DMAPP into geranyl diphosphate (GPP), which is cyclized by pinene synthase (PS) to produce pinene. Pinene can be dimerized into pinene dimers via chemical catalysis to generate a high energy density biosynthetic tactical fuel. b. Construct design. The mevalonate pathway is present in one plasmid as an operon under control of the LacUV5 promoter. The PS and GPPS are present in separate plasmids as an operon under control of the Trc promoter. Also shown is the feedback regulation at the end of the pinene biosynthetic pathway. GPP inhibits GPPS and PS, while pinene inhibits PS. c. Pinene synthase cyclization mechanism of GPP to pinene.

1.3 Results

1.3.1 Identifying pinene synthases

Our objective was to develop an *E. coli* platform for the production of pinene that, without any extra purification, could be used for the synthesis of pinene dimers.

Therefore, we focused on high fidelity pinene synthases that cyclize GPP exclusively into α- or β-pinene. For use as biofuels, we have no preference with respect to which pinene isomer to produce, as both isomers can be transformed into pinene dimers[11]. For use as commodity chemicals, however, β-pinene is more valuable, as it is less abundant than α-pinene in turpentine[15]. From the eleven pinene synthases characterized to date[31], we selected three pinene synthases with the highest fidelity (pinene synthases that produce pinene almost exclusively) for testing. Specifically, we selected *P. taeda* α-pinene synthase, which produces 100% α-pinene[32]; the mixed α/β pinene synthase from *A. grandis*, which produces ~42% α- and ~58% β-pinene[33]; and the mixed α/β pinene synthase from *P. abies*, which produces ~57% α- and ~27% β-pinene[34]. Conifer PSs, unlike sesquiterpene synthases or angiosperm PSs, prefer manganese over magnesium as a co-factor for catalysis.

1.3.2 Microbial production of pinene (co-expression)

To identify the geranyl diphosphate/pinene synthase (GPPS/PS) pair that gives rise to the highest microbial production of pinene, we combinatorially screened the suite of three PSs and three GPPSs in an *E. coli* strain able to overproduce IPP and DMAPP. To overproduce IPP and DMAPP in *E. coli*, we introduced into *E. coli* MG1655 one vector carrying the genes from the mevalonate pathway to convert acetyl-CoA into IPP and DMAPP[17]. A second vector carried the genes for GPPS and PS in an operon driven by the strong IPTG-inducible promoter P_{Trc} (Figure 1.2). Since GPPS and PS genes are endogenously expressed in plant plastids, we removed the plastid signal peptide and codon optimized their sequences to match the *E. coli* codon usage. Production of pinene

by *E. coli* transformed with all combinations of the three PSs and three GPPSs was examined (Figure 1.3**)**.

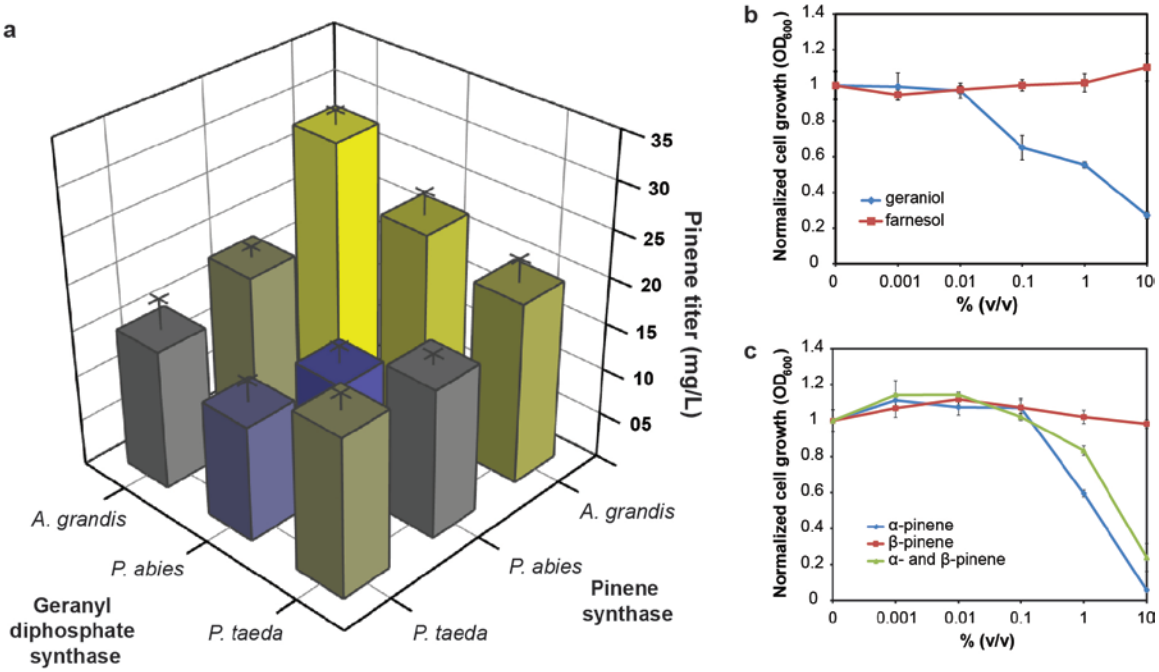

Figure 1.3: Microbial production of pinene via co-expression of geranyl diphosphate synthase and pinene synthase. a. Pinene titers as a function of the GPPS and PS co-expressed. b. Toxicity of geraniol, a proxy for GPP, and farnesol, a proxy for FPP, on cell growth. Cell growth was normalized to the growth of the same strain in the absence of geraniol or farnesol. c. Toxicity of intracellularly-produced GPP. Cell growth was normalized to the growth of the same strain not expressing GPPS. The experiments shown in a., b., and c. were done in triplicate and the error bars represent the standard deviation from the mean.

The GPPS/PS combination with the highest pinene titer was that from *A. grandis* GPPS/PS at 27.9 mg/L, while the lowest pinene titer at 11.2 mg/L resulted from the *P. abies* GPPS/PS pair. The same trend was found when comparing pinene specific production: *A. grandis* GPPS/PS led with 15.6 mg/L/OD_{600}, while *P. abies* GPPS/PS and *P. taeda* GPPS/PS had the lowest, at 3.0 mg/L/OD_{600} (Figure 1.4). The three microbial

platforms with the highest pinene titers harbored the *A. grandis* PS, independent of the source of the GPPS used.

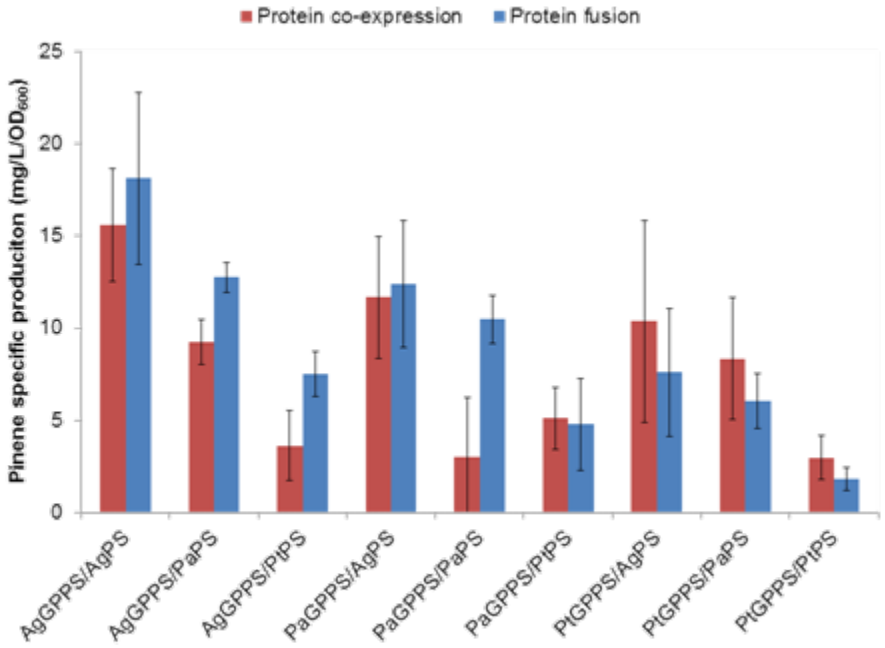

Figure 1.4: Pinene specific production using protein co-expression and protein fusions. The cultures were grown at 37°C until they reached OD_{600}=0.8 before induction with 1mM IPTG and overlayed with 20% dodecane. Pinene titers and cell growth were measured after 72hrs. Shown is the average of three experiments, the error bars represent the standard deviation from the mean.

Finally, although the cells expressing the *P. abies* GPPS/PS pair, the *P. taeda* GPPS/PS pair, and *A. grandis* GPPS/*P. taeda* PS pair grew to a higher final cell density than the other pinene-producing strains (Figure 1.5), all of the strains had similar specific growth rates (Figure 1.6).

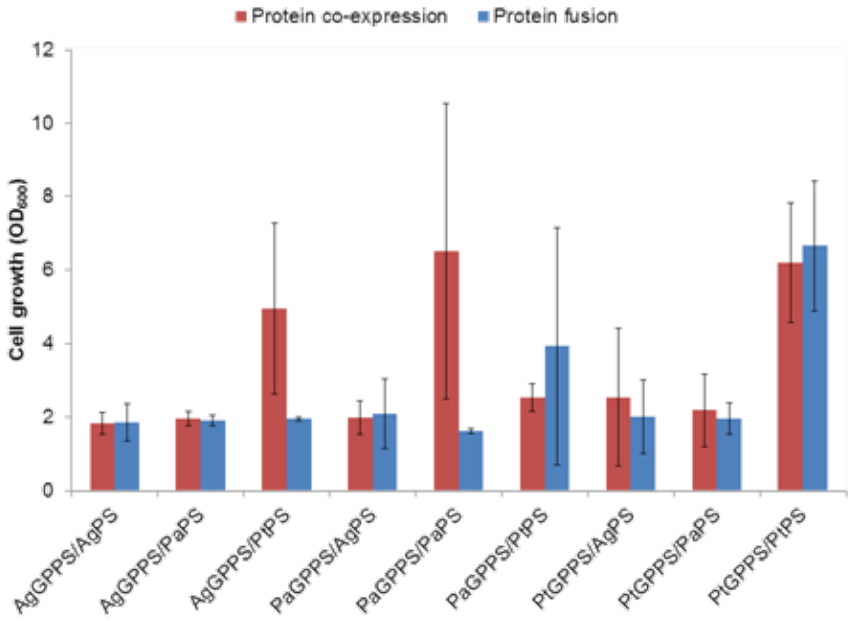

Figure 1.5: Cell growth after 72hrs of pinene microbial production. The cultures were grown at 37°C for 3 hrs before induction with 1mM IPTG and overlayed with 20% dodecane. Cell growth was measured after 72hrs. Shown is the average of three experiments and the error bars represent the standard deviation from the mean.

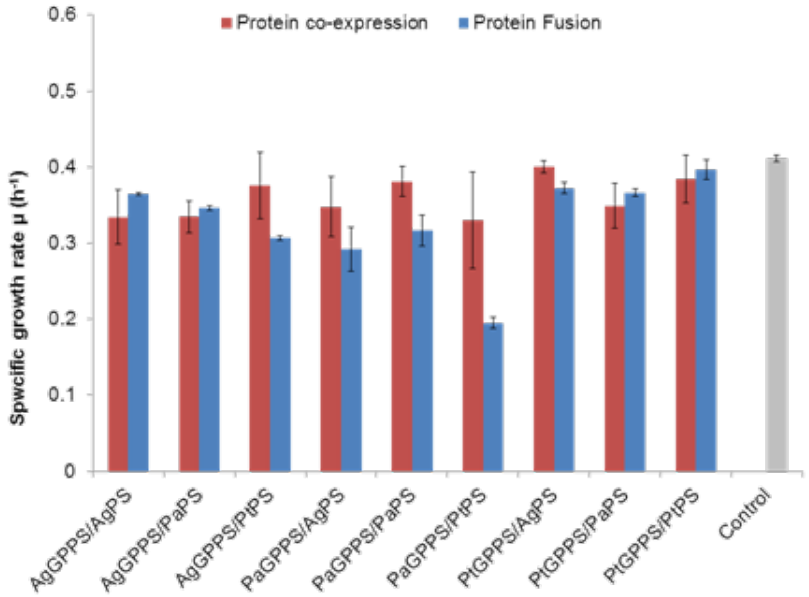

Figure 1.6: Specific growth rate of the strains used in the pinene microbial production. The cultures were grown in EZ-rich media with 1% glucose at 37°C for

3 hrs before induction with 1mM IPTG, and grown until they reached saturation (20hrs). The control is *E. coli* MG1655 harboring both an ampicillin and a chloramphenicol vector.

To investigate whether PS or GPPS protein expression limited the performance of the pinene-producing microbes, we determined the protein levels of both soluble and insoluble PS and GPPS. With respect to PS, *P. abies* PS expressed the best in both the soluble and insoluble fraction. Remarkably, *A. grandis* PS resulted in the highest pinene titers even though the proteins in the soluble and insoluble fractions were barely visible on the Western Blot (Figure 1.7). With respect to GPPS, *P. abies* and *A. grandis* GPPSs had similarly robust protein expression in the soluble and insoluble fractions, and significantly higher protein expression than *P. taeda* GPPS (Figure 1.8). Interestingly, *P. taeda* GPPS resulted in pinene titers on par with *P. abies* GPPS, hinting at higher enzymatic activity or reduced inhibition of *P. taeda* GPPS by GPP.

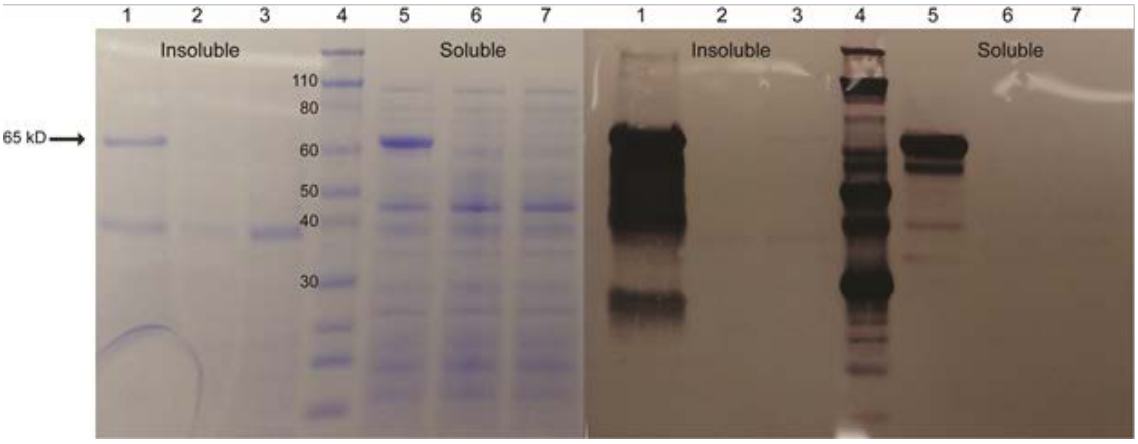

Figure 1.7: SDS-PAGE gel and western blot of pinene synthases (PSs). SDS-PAGE (4-12% BisTris) and western blot were run simultaneously with equal loading. Lanes: 1. *Picea abies* PS (65 kD) insoluble fraction, 2. *Abies grandis* PS (67 kD) insoluble fraction, 3. *Pinus taeda* PS (70 kD) insoluble fraction, 4. Ladder, 5. *Picea abies* PS soluble fraction, 6. *Abies grandis* PS soluble fraction, 7. *Pinus taeda* PS soluble fraction. Western Blot has the same line arrangement as coommasie stained SDS-PAGE gel. For westerns, anti-His-tag was purchased from Sigma (H1029) and secondary antibody, goat anti-mouse was also purchased from Sigma (A3688).

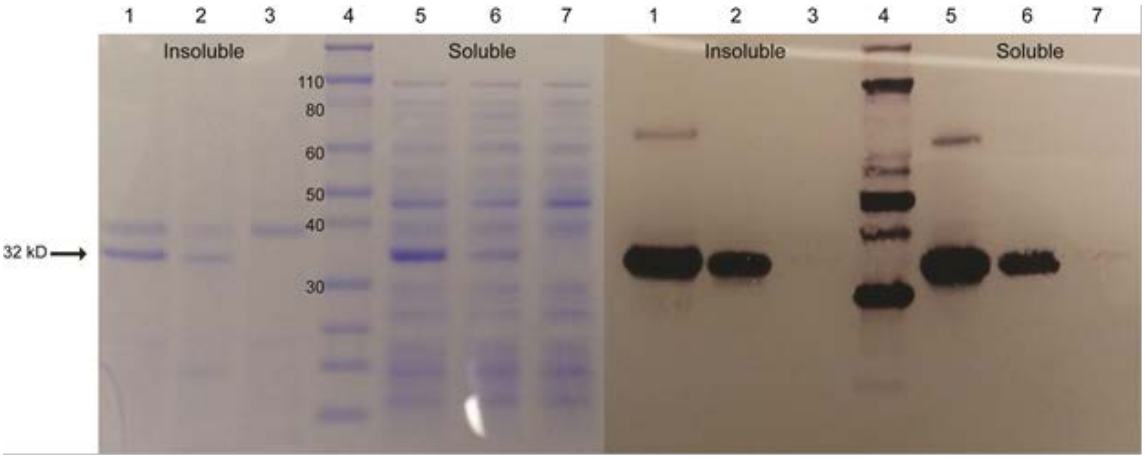

Figure 1.8: SDS-PAGE gel and western blot of geranyl diphosphate synthases (GPPSs). SDS-PAGE (4-12% BisTris) and western blot were run simultaneously with equal loading. Lanes: 1. *Picea abies* GPPS (32 kD) insoluble fraction, 2. *Abies grandis* GPPS (32 kD) insoluble fraction, 3. *Pinus taeda* GPPS (32 kD) insoluble fraction, 4. Ladder, 5. *Picea abies* GPPS soluble fraction, 6. *Abies grandis* GPPS soluble fraction, 7. *Pinus taeda* GPPS soluble fraction. Western Blot has the same line arrangement as coommasie stained SDS-PAGE gel. For westerns, anti-His-tag was purchased from Sigma (H1029) and secondary antibody, goat anti-mouse was also purchased from Sigma (A3688).

Considering this information, we conclude that, in this system, *A. grandis* PS is the most efficient PS. Further, combinatorial screening of GPPS/PS pairs enabled us to reach pinene titers of ~28 mg/L using synthetic defined medium (EZ-rich) containing 1% glucose without medium optimization, which is a 5-fold increase over the previously reported 5.4 mg/L *E. coli* production of pinene in shake flasks[18] using complex medium containing 2% glucose and which required medium optimization. It is possible that pinene production is currently limited by 1) pinene or GPP toxicity to the microbial host, which may limit cell growth and further increases in pinene titers; 2) GPP inhibition of GPPS activity, resulting in low GPP pools for PS to act upon; or 3) diversion of GPP for the production of endogenous pyrophosphates, such as farnesyl pyrophosphate.

1.3.3 Geraniol toxicity

As the same *E. coli* IPP and DMAPP overproduction strain has been previously used to produce the sesquiterpene bisabolene at ~400 mg/L[4], the levels of IPP and DMAPP are not currently limiting pinene titers. We hypothesized that high intracellular GPP concentration may be toxic to the cell by, for example, inserting itself into the cell membrane. We measured GPP toxicity to *E. coli* in two ways: 1) by exogenous addition of geraniol, a proxy for GPP, to the microbial culture; and 2) by accumulating GPP intracellularly, via overexpression of GPPS. We used geraniol as a proxy for GPP because 1) both are C10 compounds and should inflict similar disruption on the cell membrane, due to their identical tails; 2) GPP would rapidly dephosphorylate when exogenously added to the medium resulting in geraniol; and 3) GPP may also dephosphorylate intracellularly to geraniol. Addition of geraniol to the medium results in reduced cell growth starting at 0.05% v/v, which translates to 445 mg/L of geraniol (Figure 1.3). In contrast, farnesol, a proxy for the C15 diphosphate FPP, and the intermediate to bisabolene[4], shows no toxicity to *E. coli*. The concentration of geraniol required to observe reduced cell growth is currently one order of magnitude higher than the titers of microbially-produced pinene. However, it is possible that internally produced GPP is more toxic than exogenously added geraniol. To test this hypothesis, we expressed only the GPPS in the IPP and DMAPP overproduction strain with the goal of accumulating GPP intracellularly. Interestingly, the specific growth rate of the IPP and DMAPP overproduction strain, in the absence of GPPS and, expressing instead, a terpene synthase to control for the cell burden of plasmid maintenance and protein expression (μ= 0.31 h^{-1}), was similar to that of the same strain expressing *A. grandis* GPPS (μ= 0.41 h^{-1}), *P. abies* GPPS (μ= 0.38 h^{-1}), or *P. taeda* GPPS (μ= 0.30 h^{-1}).

Analysis of the cell culture revealed no geraniol produced in the strains expressing GPPS. We do note that, when measuring toxicity of intracellularly produced GPP, it is possible that, in the absence of PS, GPPS does not overproduce GPP, and, therefore, we do not see a detrimental effect of GPP. In summary, however, we cannot conclude that GPP toxicity to the cell host limits pinene production.

1.3.4 Pinene toxicity

We measured the toxicity of both α- and β-pinene via exogenous addition and found that α-pinene is more toxic than β-pinene (Figure 1.3). Specifically, α-pinene results in reduced cell growth starting at 0.5% v/v (~ 4.3 g/L) pinene in the medium. This concentration is higher than the one obtained using the best pinene-producing microbe (~ 28 mg/L). Cells harboring *P. taeda* PS, which preferentially produces α-pinene, should be the platform most highly impacted by α-pinene toxicity. However, cells harboring *P. taeda* PS had robust cell growth (Figure 1.5). Given that the two other PSs produce a mixture of α- and β-pinene, we determined the toxicity of a 50:50 mixture of α:β pinene. The α/β-pinene mixture reduced cell growth starting at 1% v/v (~ 8.6 g/L) pinene in the medium, significantly higher than that produced by our *A. grandis* GPPS-PS platform (~ 28 mg/L). Judging from these data, the titers produced by our engineered strains were not close to toxic pinene levels. However, toxicity could only be measured by exogenously added pinene, and it is possible that internally produced pinene may be more toxic to the cell.

1.3.5 Overcoming GPPS inhibition by GPP via protein fusions

To overcome potential GPPS inhibition by GPP, which may result in low GPP pools for PS to act upon, we envisioned generating protein fusions where the GPPS and PS active sites face one another so that GPP could be channeled from the GPPS active site directly into the PS active site. We used structural information to generate homology models of GPPS fused to PS in different orders, with the goal of determining the protein fusion directionality that would result in the active site of these enzymes facing one another. Specifically, using Phyre2[35], we constructed structural homology models of *A. grandis* GPPS fused to *A. grandis* PS, the GPPS/PS enzyme pair leading to the best pinene titers. The homology model used *M. piperita* GPPS large subunit[28] to thread *A. grandis* GPPS and *Mentha spicata* limonene synthase[36] to thread *A. grandis* PS. Placing GPPS at the N-terminus of the fusion resulted in the PS and GPPS active sites facing one another, while placing PS at the N-terminus resulted in the PS and GPPS active sites facing in the same direction (Figure 1.9). Therefore, we placed GPPS at the N-terminus and PS in the C-terminus of the protein fusion. To optimize the linker length between GPPS and PS, we tested linkers of three, six, and nine amino acids for the fusion (Figure 1.9). The three- and six-amino acid linkers performed slightly better than the nine-amino acid linker. To ensure enough space for pinene release from the PS active site at the end of the reaction, we chose the six-amino acid linker length. To experimentally confirm that the directionality of the protein fusion was important, we constructed a protein fusion with PS at the N-terminus and GPPS at the C-terminus using the same six-amino acid linker. Constructing the protein fusion in this manner, with the active sites facing the same direction rather than each other, resulted in a pinene titer of 4 mg/L, which was even lower than the pinene titer obtained via co-expression of the same

PS and GPPS. With this result in hand, we proceeded to test only fusions with the PS and GPPS active sites facing one another for increased activity.

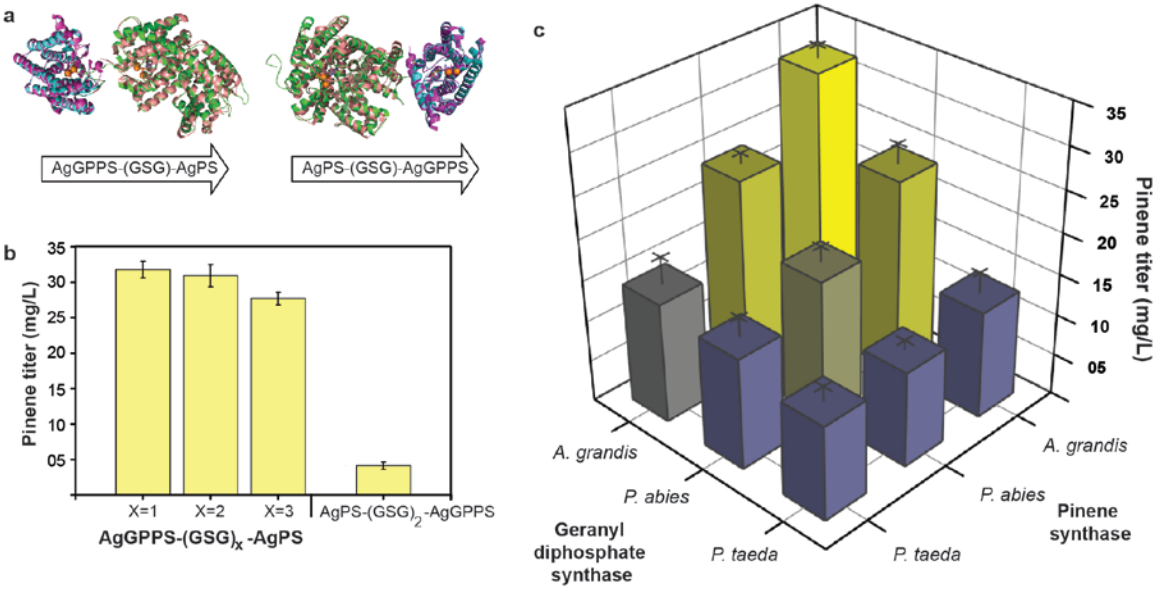

Figure 1.9: Microbial production of pinene via fusion of geranyl diphosphate synthase and pinene synthase. a. Fusion protein design. GPPS at the N-terminus of the fusion results in GPPS and PS active sites facing one another. PS at the N-terminus of the fusion results in GPPS and PS facing in the same direction. Peach: *M. spicata* limonene synthase structure (PDB: 2ONG). Magenta: *M. pipetira* GPPS (PDB: 3OAC). Green and cyan: PS fused to GPPS. b. Pinene titers of protein fusions with three, six, and nine amino acid linker lengths. c. Microbial pinene titers as a function of the GPPS and PS protein fusions. The experiments shown in b. and c. were done in triplicate, and the error bars represent the standard deviation from the mean.

1.3.6 Pinene microbial production using protein fusions

Not knowing if *A. grandis* GPPS-$(GSG)_2$-PS would result in the highest pinene production, we constructed all possible combinations of GPPS-$(GSG)_2$-PS protein fusions, expressed them under the strong Trc promoter, and analyzed the resulting pinene production (Figure 1.9). *A. grandis* GPPS-$(GSG)_2$-PS resulted in the highest pinene titer, at 32.4 mg/L, while *P. taeda* GPPS-$(GSG)_2$-PS resulted in the lowest pinene titer, at 11.4

mg/L. The *A. grandis* GPPS-PS fusion pair produced the most pinene in the co-expression (non-fusion) experiments as well. Overall, only four of the nine protein fusions showed improvement in pinene titers when compared to co-expression. The *P. abies* GPPS-PS fusion pair had the greatest improvement in pinene production, 52% when compared to co-expression, followed by the *A. grandis* GPPS-*P. abies* PS fusion, with a 35% improvement. The majority of constructs showed decreased pinene titers with the most significant drop in production coming from the *P. taeda* GPPS-*A. grandis* PS fusion, which dropped by 34%. The *A. grandis* GPPS-PS fusion pair led in the specific production of pinene, with 18.1 mg/L/OD_{600}, while the *P. taeda* GPPS-PS fusion pair had the lowest, at 1.8 mg/L/OD_{600}. (Figure 1.4).

All GPPSs and PSs investigated in this study have the same conifer origin: *A. grandis*, *P. abies* and *P. taeda*. We hypothesized that fusing enzymes from the same conifer species might result in higher pinene production than co-expressing them. Consistent with this hypothesis, we found that fusing the GPPS and PS from the same species in the case of *A. grandis* resulted in increased production over the co-expression. However, matching the GPPS and PS from the same species in the case of *P. taeda* worsened production.

1.3.7 Rationalization of operon vs. protein fusion performance

Previous protein fusions of FPPS to C_{15} sesquiterpene synthases (bisabolene synthase and farnesene synthase) yielded 2- to 8-fold improvements in sesquiterpene titers[37-38]. Protein fusions of geranylgeranyl diphosphate synthase to diacylglycerol diphosphate phosphatase resulted in ~2.5-fold improvement geranylgeraniol titers[39]. The marginal

improvements seen in pinene titers when fusing GPPS to PS can be rationalized by the inhibition of PS by GPP. It is possible that, by bringing GPPS and PS together, the inhibition of GPPS by GPP is overcome, but the inhibition of PS by GPP is exacerbated, resulting in a marginal overall improvement of pinene titers. To address this hypothesis, we determined the activity of PS in the presence of both manganese and magnesium in cell lysate at different GPP concentrations and confirmed that pinene synthesis decreases with increasing GPP concentration (Figure 1.10). Percent conversion of GPP to pinene by *A. grandis* PS decreased from 14.5% to 2.9% over the 0.1-1 mM range tested when manganese, the preferred co-factor, was used. The percent conversion decreased less sharply, from 11% to 5.4% over the same range, when using magnesium, which is the most likely co-factor used in *E. coli*. Next, we measured the activity of PS when it was part of the GPPS-PS protein fusion, in order to determine if the inhibition of PS was exacerbated when part of the fusion. In the presence of manganese, the *A. grandis* PS alone is 23% more inhibited than the *A. grandis* GPPS-PS protein fusion (Figure 1.10). In the presence of magnesium, however, the *A. grandis* GPPS-PS protein fusion is 33% more inhibited than the *A. grandis* PS alone (Figure 1.10). Given that magnesium is the most likely cofactor in *E. coli*, we conclude that inhibition of PS is exacerbated when it is part of a fusion in our system.

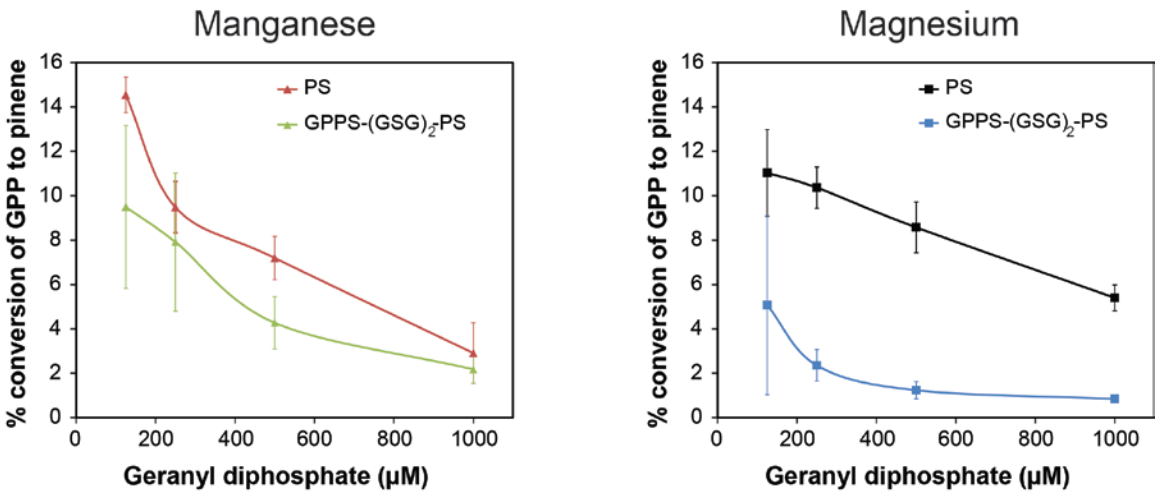

Figure 1.10: Analysis of GPP inhibition of pinene synthase alone and when part of a protein fusion with geranyl diphosphate synthase. A. Percent conversion of GPP to pinene in cell extract of *E. coli* expressing the *A. grandis* PS (PS) alone or the *A grandis* geranyl diphosphate synthase-pinene synthase (AgGPPS-(GSG)$_2$-AgPS) protein fusion in the presence of manganese. B. Percent conversion of GPP to pinene in cell extract of *E. coli* expressing the *A. grandis* PS (PS) alone or the *A grandis* geranyl diphosphate synthase-pinene synthase (AgGPPS-(GSG)$_2$-AgPS) protein fusion in the presence of magnesium. Percent conversion was calculated by dividing the moles of pinene produced over the moles of GPP at the beginning of the reaction. The experiments were done in triplicate, and the error bars represent the standard deviation from the mean.

1.3.8 Ratios of α- and β-pinene

Knowing that α-pinene was at least potentially toxic at high titers, and noting that β-pinene is more expensive than α-pinene as it is less common in turpentine[15], we investigated the α:β isomer ratios produced by the pinene-producing microbes. From the three pinene synthases, only *P. taeda* PS produced primarily α-pinene while *A. grandis* PS and *P. abies* PS produced α/β-pinene mixtures. In the heterologous production of pinene, we expected that only the identity of the PS would determine the ratio of pinene isomers. However, we found that the identity of the GPPS, and whether or not the GPPS/PS pair was a fusion, affected the ratio of pinene isomers (Figure 1.11). *P. taeda*

PS consistently produced more α-pinene than β-pinene, independent of which GPPS it was paired with, or whether it was co-expressed or part of a protein fusion. In contrast, the pinene isomer profile of *P. abies* PS and *A. grandis* PS changed depending on the identity of the GPPS with which they were paired. *A. grandis* PS resulted in a ~50:50 mixture of α- to β-pinene when paired with *A. grandis* GPPS. However, when paired with *P. taeda* GPPS, we observed a significant increase in α-pinene production, whether the enzymes were co-expressed or fused. The case with *P. abies* PS is more dramatic. Pairing *P. abies* PS with *P. taeda* GPPS resulted in mostly α-pinene production, while pairing it with *P. abies* GPPS as a co-expression resulted in mostly β-pinene and, as a fusion, resulted in a significant reduction of α-pinene.

In terpene synthases, the N-terminal strand caps the active site, shielding the terpenyl carbocation from water[36,40]. Specifically, in monoterpene synthases, the N-terminus contains the $RR(X)_8W$ motif, where a pair of arginines stabilizes the holo conformation of the protein[36]. The pair of arginines has also been implicated in the isomerization of GPP to linalyl diphosphate, which is necessary to reach the α-terpenyl cation and, in turn, the pinyl carbocation and the final pinene structure[41] (Figure 1.2). In the protein co-expression experiments, the PS cyclizes GPP into pinene without any movement limitation to the N-terminal strand; therefore, we do not expect any change in the PS catalytic mechanism. Rather, the only difference between the different co-expressions should be the level of GPP produced by the different GPPSs, and thus available to the PS for catalysis. The GPP levels are determined by the origin of the GPPS (i.e., *A. grandis*, *P. taeda*, *P. abies*) with which the PS is co-expressed. Given the observed differences in α:β pinene ratios when *P. abies* PS and *A. grandis* PS are coupled to different GPPS, we

hypothesize that GPP concentration regulates the product specificity of *P. abies* PS and *A. grandis* PS. To address this hypothesis, we measured the pinene isomer ratio at different GPP concentrations using the cell extract of *E. coli* expressing *A. grandis* PS, the PS with the greatest change in pinene isomer ratio *in vivo* (Figure 1.11). Increasing the GPP concentration decreases the α:β pinene ratio from 1.76 at 1.25 μM GPP to 1.30 at 1 mM GPP when manganese is used as the *A. grandis* PS co-factor. Although manganese is the preferred *A. grandis* PS cofactor, the *E. coli* cytosol has at has at least 100-fold higher concentration of magnesium than manganese[42]; thus, we also tested the α:β pinene ratio in the presence of magnesium. Using magnesium, we also see a similar decrease in the α:β pinene ratio from 1.30 at 1.25 μM GPP to 0.98 at 1mM GPP. PS isomer ratio regulation by GPPS may have been previously overlooked, as the PS product profile has been only measured at a single GPP concentration, and kinetic characterization of PS is traditionally carried out using radioactive or colorimetric assays.

In the protein fusion experiments, the movement of the PS N-terminal strand is restricted by its fusion to GPPS, which, among other things, may cause improper capping of the active site to water (Figure 1.11). Additionally, the proximity of the N-terminal strand to the J-K loop, which stabilizes the holo form of the enzyme[40], may explain the change in the pinene isomer ratio seen in the fusion proteins. Residues in the J-K loop have been previously shown to alter the terpene cyclization mechanism[43]. Specifically, a change in the pinene isomer ratio of *A. grandis* PS has been achieved via a single mutation at the end of the J-helix (F597W)[44]. We hypothesize that the limited movement of the N-terminal strand hinders the proper localization of the J-K loop and, in turn, the J-helix, thus changing the pinene isomer ratios.

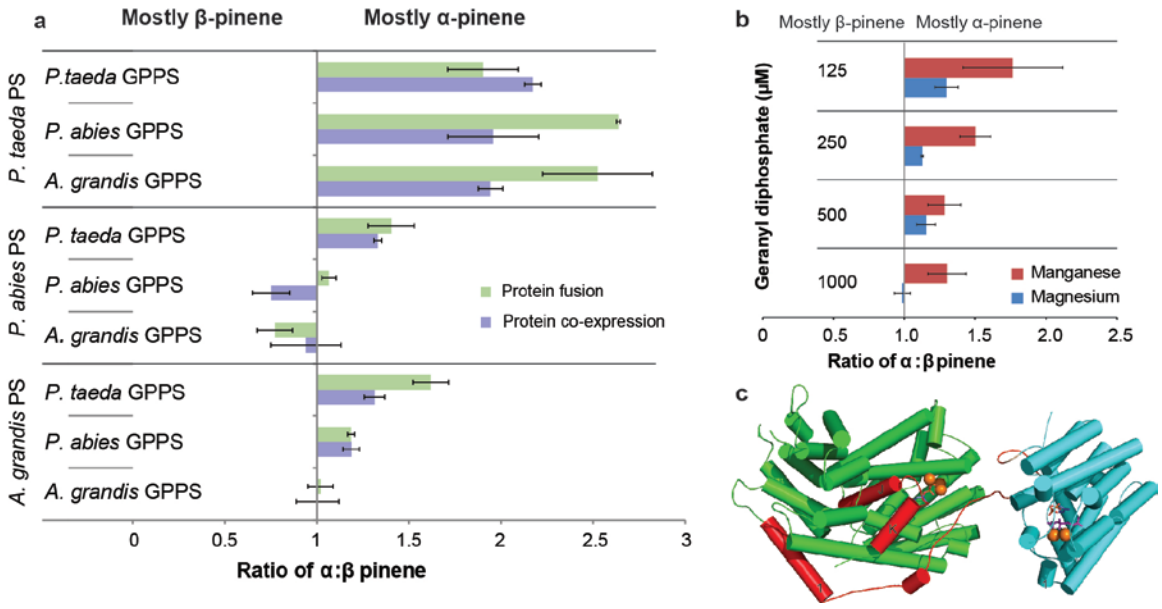

Figure 1.11: Microbially-produced pinene isomer ratios as a function of geranyl diphosphate synthase (GPPS) and pinene synthase (PS). a. Ratio of α:β-pinene when matching different GPPSs and PSs using either operons of fusions. b. Ratio of α:β-pinene in cell extract of *E. coli* expressing *A. grandis* PS with either manganese or magnesium as the cofactor. c. Homology model of *A. grandis* GPPS fused to *A. grandis* PS. The N-terminal strand and the F-K helix and loop of GPPS are highlighted in red. The experiments shown in a. and b. were done in triplicate, and the error bars represent the standard deviation from the mean.

1.4 Discussion

We have engineered *E. coli* for the renewable production of pinene, the immediate precursor to pinene dimers, a biosynthetic alternative to JP-10. By testing PSs and GPPSs from several organisms, we increased pinene production to 27.9 mg/L, two-fold higher than the lowest pinene producing pair (11.2 mg/L). Combinatorial protein fusions of GPPS and PS resulted in slight improvement in pinene titers to ~32 mg/L using *A. grandis* GPPS-$(GSG)_2$-PS. We rationalize the marginal improvement of pinene titers using protein fusions as due to PS inhibition by GPP, counterbalancing any improvements obtained by relief of GPPS inhibition by GPP. Therefore, we conclude

that enzyme inhibition patterns must be taken into account to help determine the potential success or failure of a protein fusion strategy. Given that the IPP- and DMAPP-overproducing *E. coli* strain used in this study has been previously used to achieve ~400 mg/L of bisabolene, IPP and DMAPP precursor levels are not limiting; hence, the current bottleneck must lie in the last two enzymes in the pinene biosynthetic pathway. Therefore, we believe that further improvements in pinene titer will come from a combination of protein engineering approaches, such as generation of PSs with decreased GPPS inhibition, and better control of GPP intracellular levels.

Using Flux Balance Analysis[45], we calculated the theoretical yield of pinene produced from *E. coli* using the mevalonate pathway to be 0.270 g pinene/g glucose. Experimentally, we produced pinene at ~1.2% of the pathway-dependent calculated theoretical yield, that is, assuming only glucose in the EZ-rich media is used for pinene production. Assuming a break-even price of glucose at the mill to be close to US $0.10/lb[4], we calculate that the raw material cost of pinene production, ignoring non-sugar costs, to be approximately $68/kg of pinene. Assuming raw material costs to be only 50% of the final cost[46], the final price of pinene would be $136/kg, or $443/gal, at current production levels. Assuming that we could produce pinene at the theoretical yield, the final price would be $1.63/kg, or $5.31/gal of pinene, resulting in a final price of ~$6.42/gal for pinene dimers (assuming 90% yield[11] and negligible conversion cost), a significant savings over the current price of JP-10 ~$25/gal[47]. Commercial viability of microbial pinene for use as tactical fuel therefore requires reaching 26% theoretical yield.

Analysis of the pinene isomer profile of our microbial strain revealed that the ratios of α- to β-pinene vary not only due to the identity of the PS, but also due to the identity of

GPPS and whether the proteins are co-expressed or in a fusion. We demonstrated that the concentration of GPP available to *A. grandis* PS for catalysis alters the pinene isomer ratios, which explains the changes in isomer ratio seen in the protein co-expression experiments. A crystal structure of PS would aid in elucidating the mechanism of how GPP affects pinene synthase isomer ratio. Modulation of PS activity with GPP concentration may also occur *in planta* in response to stress. Indeed, changes in the α-pinene to β-pinene ratios have been previously seen after methyl jasmonate (MeJa) induction in pine[48]. We hypothesize that, in the GPPS-PS protein fusions, conformational restriction of the PS N-terminal strand may alter the holo conformation of PS, specifically the J-K loop, thus altering the terpene cyclization mechanism. *In vitro* biochemical analysis using purified PS-GPPS enzyme will be required to confirm the proposed changes in terpene cyclization mechanism.

Future engineering of microbial pinene production will require addressing the pathway problems at the enzyme level. Given that *A. grandis* PS resulted in the highest pinene production even though it was one of the more poorly expressed PSs, improving expression of this enzyme should result in increased pinene production. Alternatively, identifying GPPS and PS mutants not inhibited by GPP may also aid in increasing pinene titers. Finally, all the PSs in this study prefer manganese as a co-factor over magnesium, which is far more available in the *E. coli* cytosol, and, therefore, necessarily utilized in our experiments. Utilization of the magnesium may be limiting the activity of the PS *in vivo*; thus, identifying and expressing magnesium-dependent PSs may yield higher pinene titers as well.

1.5 Methods

1.5.1 Plasmid construction

1.5.1.1 Construction of ptXPS

To construct ptAgPS, AgPS truncated at $RR(X)_8W$ to remove the plasmid signal peptide and was amplified from pSynAgPS using primers PPY782/PPY969. To construct ptPaPS, PaPS truncated at $RR(X)_8W$ and was amplified from pSynPaPS using primers PPY1220/PPY1212. To construct ptPtPS, PtPS truncated at $RR(X)_8W$ and was amplified from pSynPtPS using primers PPY642/PPY643. All inserts were cloned into pTRC99 at NcoI/XmaI.

1.5.1.2 Construction of pXGPPS

To construct pAgGPPS, AgGPPS was amplified from pSynAgGPPS, signal peptide removed, using primers PPY608/PPY630. To construct pPaGPPS, PaGPPS was amplified from pSynPaGPPS, signal peptide removed, using primers PPY1210/PPY1227. To construct pPtGPPS, was amplified from pSynPtGPPS, signal peptide removed, using primers PPY1219/1225. All inserts were cloned into pTRC99 at NcoI/XmaI.

1.5.1.3 Linker length optimization constructs

To construct AgGPPS-$(GSG)_1$-AgPS, $(GSG)_1$-AgPS was amplified from pAgPS using primers SS90/SS87 and cloned into pAgGPPS at XmaI/HindIII. Primer SS90 introduces the $(GSG)_1$ sequence. To construct AgGPPS-$(GSG)_3$-AgPS, $(GSG)_3$-AgPS was amplified from pAgGPPS-$(GSG)_2$-AgPS using primers SS89/SS87 and cloned into pAgGPPS at XmaI/HindIII. Primer SS89 introduces and additional GSG sequence to construct $(GSG)_3$. To construct pAgPS-$(GSG)_2$-AgGPPS, AgGPPS was amplified from

pAgGPPS using primers SS117/SS118 and cloned into pAgPS at XmaI/HindIII. Primer SS117 introduces the $(GSG)_2$ linker.

1.5.1.4 His6-tagged PS and GPPS constructs

$(His)_6$-AgGPPS: AgGPPS was amplified from pAgGPPS using primers SS70/SS76. $(His)_6$-PaGPPS: PaGPPS was amplified from pPaGPPS using primers SS72/SS76. $(His)_6$-PtGPPS: PtGPPS was amplified from vector pPtGPPS using primers SS71/SS76. $(His)_6$-AgPS: AgPS sequence was amplified from pAgPS using primers SS79/SS76. $(His)_6$-PaPS: PaPS was amplified from pPaPS using primers SS75/SS76. $(His)_6$-PtPS: PtPS was amplified from pPtPS using primers SS74/SS76. All inserts were cloned using SLIC to pTrc99 amplified using primers SS77/SS78.

1.5.1.5 His6-tagged GPPS-PS protein fusion constructs

$(His)_6$-AgGPPS-$(GSG)_2$-AgPS: AgGPPS-$(GSG)_2$-AgPS was amplified from pAgGPPS-$(GSG)_2$-AgPS using primers SS169/SS170. $(His)_6$-PaGPPS-$(GSG)_2$-AgPS: PaGPPS-$(GSG)_2$-AgPS was amplified from pPaGPPS-$(GSG)_2$-AgPS using primers SS171/SS170. $(His)_6$-PtGPPS-$(GSG)_2$-AgPS: PtGPPS-$(GSG)_2$-AgPS was amplified from vector pPtGPPS-$(GSG)_2$-AgPS using primers SS172/SS170. $(His)_6$-PaGPPS-$(GSG)_2$-PaPS: PaGPPS-$(GSG)_2$-PaPS sequence was amplified from pPaGPPS-$(GSG)_2$-AgPS using primers SS171/SS170. $(His)_6$-PtGPPS-$(GSG)_2$-PaPS: PtGPPS-$(GSG)_2$-PaPS was amplified from PtGPPS-$(GSG)_2$-PaPS using primers SS172/SS170. $(His)_6$-AgGPPS-$(GSG)_2$-PaPS: AgGPPS-$(GSG)_2$-PaPS was amplified from pAgGPPS-$(GSG)_2$-PaPS using primers SS169/SS170. $(His)_6$-PtGPPS-$(GSG)_2$-PtPS: PtGPPS-$(GSG)_2$-PtPS was amplified from pPtGPPS-$(GSG)_2$-PtPS using primers SS172/SS170. $(His)_6$-PaGPPS-

$(GSG)_2$-PtPS: PaGPPS-$(GSG)_2$-PtPS was amplified from pPaGPPS-$(GSG)_2$-PtPS using primers SS171/SS170. $(His)_6$-AgGPPS-$(GSG)_2$-PtPS: AgGPPS-$(GSG)_2$-PtPS was amplified from pAgGPPS-$(GSG)_2$-PtPS using primers SS169/SS170. All inserts were cloned using SLIC to pAgGPPS-$(GSG)_2$-AgPS digested with NcoI and HindIII.

1.5.1.6 Construction of pPS-GPPS (coexpressions)

The nine PS and GPPS genes were codon optimized to match the *E. coli* codon usage and commercially synthesized (Genescript). For the pAgPS-XGPPS series, AgGPPS, PaGPPS, and PtGPPS were amplified from pAgGPPS, pPaGPPS, and pPtGPPS, respectively using primers PPY61/PPY59 and cloned into pAgPS at BamHI/HindIII. For the pPaPS-XGPPS series, AgGPPS, PaGPPS and PtGPPS were amplified from pAgGPPS, pPaGPPS and pPtGPPS, respectively using primers PPY60/PPY59 and cloned into pPaPS at BamHI/HindIII. For the pPtPS-XGPPS series, AgGPPS, PaGPPS and PtGPPS were amplified from pAgGPPS, pPaGPPS and pPtGPPS, respectively using primers PPY58/PPY59 and cloned into pPtPS at BamHI/HindIII.

1.5.1.7 Construction of pGPPS-$(GSG)_2$-PS (protein fusions)

pAgGPPS-$(GSG)_2$-XPS series. To construct pAgGPPS-$(GSG)_2$-AgPS, AgGPPS was amplified from pAgGPPS using primers PPY90/BW33, and AgPS was amplified from ptAgPS using primers BW34/PPY91. Primer BW33 introduces the $(GSG)_2$ linker sequence. AgGPPS-$(GSG)_2$ and AgPS were fused using sawing PCR and cloned in pTRC99 at NcoI/HindIII. To construct pAgGPPS-$(GSG)_2$-PaPS, $(GSG)_2$-PaPS was amplified from pPaGPPS-$(GSG)_2$-PaPS using primers SS61/SS62 and cloned into pAgGPPS at XmaI/HindIII. To construct pAgGPPS-$(GSG)_2$-PtPS, PtPS was amplified

from pPtPS using primers SS68/SS59 primers and cloned into pAgGPPS at XmaI/HindIII. Primer SS68 introduces the $(GSG)_2$ linker sequence. pPaGPPS-$(GSG)_2$-XPS series. To construct pPaGPPS-$(GSG)_2$-AgPS, $(GSG)_2$-AgPS was amplified from pAgGPPS-$(GSG)_2$-AgPS using primers SS45/SS46 and cloned into pPaGPPS at XmaI/HindIII . To construct pPaGPPS-$(GSG)_2$-PaPS, the PaPS leader sequence was amplified from pLead using primers SS35/SS36, and the truncated PaPS sequence was amplified from ptPaPS using primers SS37/SS38. Using SLIC[49] these two pieces were cloned into pPaGPPS at XmaI/HindIII. To construct pPaGPPS-$(GSG)_2$-PtPS, PtPS was amplified from pPtPS using primers SS69/SS59 primers and cloned into pPaGPPS at XmaI/HindIII. Primer SS69 introduces the $(GSG)_2$ linker sequence. pPtGPPS-$(GSG)_2$-XPS series. To construct pPtGPPS-$(GSG)_2$-AgPS, $(GSG)_2$-AgPS was amplified from pAgGPPS-$(GSG)_2$-AgPS using primers SS49/SS50 and cloned into pPtGPPS at XmaI/HindIII. To construct pPtGPPS-$(GSG)_2$-PaPS, $(GSG)_2$-PaPS was amplified from pPaGPPS-$(GSG)_2$-PaPS using primers SS63/SS62 and cloned into pPtGPPS at XmaI/HindIII. To constuct pPtGPPS-$(GSG)_2$-PtPS, the PtPS leader sequence was amplified from pLead using primers SS31/SS32, and the truncated PtPS sequence was amplified from ptPtPS using primers SS33/SS34. Using SLIC these two pieces were cloned into pPtGPPS at XmaI/HindIII.

1.5.2 Pinene production and quantification

E. coli MG1655 was co-transformed with pBbA5c-MevT-MBI containing the mevalonate pathway and plasmids containing GPPS and PS enzymes as operons or enzyme fusions. Cultures of strains co-transformed with both plasmids were grown in LB media overnight. 100 μL of the overnight culture was used to inoculate 5 mL of EZ

rich medium (Teknova, 1% (v/v) glucose, amp^{100}, chl^{50}). The EZ rich cultures were then incubated at 37°C (250 r.p.m.) until an OD_{600} of 0.8 was reached. Then, the cultures were induced with 1 mM IPTG and overlaid with 20% dodecane. After inoculation, the cultures were incubated at 30°C (250 r.p.m.) for 72 hrs. 500 µL of the dodecane layer was placed in a 1.5-mL microcentrifuge tube, centrifuged at 25,000 x g for 1 min, and 50 µL of dodecane was diluted in 450 µL of ethyl acetate spiked with the internal standard limonene (10 µg/L). The samples were analyzed by GC-MS by using a standard curve of α (Sigma Aldrich) and (-)-β-pinene (Sigma Aldrich). The GC-MS (Agilent 7890A with Agilent 5975 MS detector) was used with an Agilent DB-5MS column. The inlet temperature was set to 300°C, flow at 1 mL/min, the oven at 50°C for 30 sec, ramp at 25°C/min to 150°C, and ramp of 40°C/min to 250°C.

1.5.3 Pinene and geraniol toxicity measurements

An overnight culture of *E. coli* strain MG1655 was used to inoculate 5 mL of EZ rich media (Teknova, 1% (v/v) glucose) containing varying concentrations of α, β, α+β-pinene, geraniol or farnesol. After inoculation, the cultures were grown at 30°C (250 r.p.m) for 24 hrs and cell growth was measured as absorption at OD_{600}.

1.5.4 Pinene conversion using cell lysate

An overnight culture of *E. coli* MG1655 transformed with pAgPS or pAgGPPS-$(GSG)_2$-PS was used to inoculate 185 mL LB medium at an OD_{600}= 0.10 and incubated at 37°C (250 r.p.m.) until an OD_{600}= 0.80 was reached. The culture was then induced with 1 mM IPTG and incubated at 30°C (250 r.p.m.) for 3 hours. The culture was then split

into two 90 mL cultures and centrifuged at 7354 x g for 20 min at 4ºC. Each pellet was resuspended in 9 mL of monoterpene synthase buffer[33] (50 mM Tris/HCl (pH= 7.5), 500 mM KCl, 5 mM dithiothreitol, 0.05% (w/v) $NaHSO_3$, and 10% (v/v) glycerol) with either 1mM $MgCl_2$ or 1mM $MnCl_2$, and sonicated using a Misonix Sonicator 3000 at 7.0 output level for 10 sec, 30 sec rest, for a total 4 min. After sonication, the lysate was centrifuged at 3220 x g for 30 min. 980μL of the supernatant was mixed with either 125 μM, 250 μM, 500 μM, or 1 mM GPP (Echelon) to a final volume of 1 mL, layered with 200 μL dodecane, and incubated at 30ºC for 2 hours. Pinene was quantified as described previously.

1.5.5 Flux balance analysis yield calculations

Contributed by Hector Garcia Martin

We used Flux Balance Analysis to determine the maximum yield for the mevalonate pathways for two different *E. coli* models (to provide extra robustness): iJR904[50] and iAF1260[51]. Using the COBRA package[52,53] we loaded each model and added the following reactions

Pt30 [c] : grdp <==> appnn + ppi (pinene synthesis)

EX_appnn(e): appnn[c] -> (exchange reaction)

HMGCOAS [c] : coa + h + hmgcoa <==> aacoa + accoa + h2o (mevalonate pathway)

HMGCOAR [c] : coa + mev-R + 2 nadp <==> 2 h + hmgcoa + 2 nadph

MEVK1 [c] : atp + mev-R --> 5pmev + adp + h

PMEVK [c] : 5pmev + atp --> 5dpmev + adp

DPMVD [c] : 5dpmev + atp --> adp + co2 + ipdp + pi

We then knocked out DXPS (setting both lower bound and upper bound to zero), set glucose input (EX_glc(e)) to -6 mMol/gdw/h, changed the objective function to maximize EX_appnn(e), and solved for the maximum allowable value. The results for both models were multiplied by 136.23/180.16 (molecular weights of pinene and glucose respectively) and divided by the incoming flux in order to obtain a gram of α-pinene per gram of glucose yield: 0.270 (iJR904) and 0.279 (iAF1266) g pinene/g glucose.

1.5.6 Protein quantification protocol

E. coli MG1655 harboring pBbA5c-MevT-MBI and one of the $(His)_6$-tagged GPPS, $(His)_6$-tagged PS, or $(His)_6$-tagged GPPS-PS were grown overnight in LB media. Overnight cultures were then inoculated 1:50 in EZ-rich media (Teknova, 1% (v/v) glucose, Amp^{100}, Chl^{30}). These cultures were grown at 37ºC, 250 r.p.m. until they reached an OD_{600} of 0.6-0.8. The cells were then induced with 500 μM IPTG and incubated at 30 ºC for 3 hours. After the 3 hour incubation, 1 mL of sample was removed from the culture medium and centrifuged at 14,000 r.p.m. for 2 min. The supernatant was discarded and the pellet was resuspended in 200 μL of phosphate buffer (Teknova). The cell suspension was sonicated three times for 20 seconds each and centrifuged at 14,000 r.p.m. for 2 min. For the soluble fraction, to 50 μL of the supernatant, 10μL of SDS loading dye was added. For the insoluble fraction, the pellet was resuspended in 50 μL of 6X SDS loading dye. Both soluble and insoluble fractions were heated at 95ºC for 15 min and then loaded onto two SDS-PAGE (20μL of each soluble and insoluble fractions).

The gels were run for 40 min at 200 V at 4 °C. After the run, one SDS-PAGE gel was stained with Coomassie Blue and the other was used for Western Blot staining.

1.6 References

1 Peralta-Yahya, P. P., Zhang, F., del Cardayre, S. B., & Keasling, J. D. Microbial engineering for the production of advanced biofuels, *Nature* **488**, 320-328 (2012).

2 Bond-Watts, B. B., Bellerose, R. J., & Chang, M. C. Y. Enzyme mechanism as a kinetic control element for designing synthetic biofuel pathways, *Nat Chem Biol* **7**, 222-227 (2011) .

3 Steen, E. J., et al. Microbial production of fatty-acid-derived fuels and chemicals from plant biomass, *Nature* **463**, 559-562 (2010).

4 Peralta-Yahya, P. P., et al. Identification and microbial production of a terpene-based advanced biofuel, *Nat commun* **2**, 483 (2011).

5 Renniger, N. & McPhee, D. Fuel compositions comprising farnesane and farnesane derivatives and method of making and using same. Patent: US20080098645 A1.

6 Braukus, M. NASA Begins Flight Research Campaign Using Alternate Jet Fuel (http://www.nasa.gov/home/hqnews/2013/mar/HQ_13-066_ACCESS.html#.UjRmpD-bGk8) (Accessed, September 2013).

7 Trimbur, D., et al. Renewable chemicals and fuels from oleginous yeast. Patent: US20110190522 A1.

8 American Society for Testing and Materials , D7566-13 Standard Specification for Aviation Turbine Fuel Containing Synthesized Hydrocarbons (2013).

9 Wright, M., Harvey, B., & Quintana, R. Highly efficient zirconium-catalyzed batch conversion of 1-butene: A new route to jet fuels, *Energ Fuel* **22**, 3299-3302 (2008).

10 Peters, M. & Taylor, J. Renewable jet fuel blendstock from isobutanol. Patent: WO20111405610.

11 Harvey, B., Benjamin, G., Wright, M., & Quintana, R. High-Density Renewable Fuels Based on the Selective Dimerization of Pinenes, *Energ Fuel* **24**, 267-273 (2010).

12 Kirby, J., & Keasling, J. D. Biosynthesis of plant isoprenoids: perspectives for microbial engineering, *Annu Rev Plant Biol* **60**, 335-355 (2009).

13 Seigler, D. Plant secondary metabolism, 2nd ed. Kluwer Academic Publisher, Norwell (1998) .

14 Breitmaier, E. Terpenes: Flavors, Frangances, Pharmaca, Pheromones, Wiley-VCH, Tubingen (2006) .

15 New Zealand Institute of Chemistry, Turpentine production and processing. (http://nzic.org.nz/ChemProcesses/forestry/4F.pdf) (2012) (Accessed November 2013).

16 Martin, V. J., Pitera, D. J., Withers, S. T., Newman, J. D., & Keasling, J. D. Engineering a mevalonate pathway in Escherichia coli for production of terpenoids, *Nat Biotechnol* **21**, 796-802 (2003) .

17 Bokinsky, G., et al. Synthesis of three advanced biofuels from ionic liquid-pretreated switchgrass using engineered Escherichia coli, *Proc Nat Acad Sci USA* **108**, 19949-19954 (2011) .

18 Yang, J. M., et al. Metabolic engineering of Escherichia coli for the biosynthesis of alpha-pinene, *Biotechnol Biofuels* **6** (2013).

19 Ajikumar, P. K., et al. Isoprenoid pathway optimization for Taxol precursor overproduction in Escherichia coli, *Science* **330**, 70-74 (2010) .

20 Alonso-Gutierrez, J., et al. Metabolic engineering of Escherichia coli for limonene and perillyl alcohol production, *Metab Eng* **19**, 33-41 (2013).

21 Dunlop, M. J., et al. Engineering microbial biofuel tolerance and export using efflux pumps, *Mol Syst Biol* **7**, 487 (2011).

22 Clastre, M., Bantignies, B., Feron, G., Soler, E., & Ambid, C. Purification and Characterization of Geranyl Diphosphate Synthase from Vitis vinifera L. cv Muscat de Frontignan Cell Cultures, *Plant Physiol* **102**, 205-211 (1993).

23 Burke, C. & Croteau, R. Geranyl diphosphate synthase from Abies grandis: cDNA isolation, functional expression, and characterization. *Arch Biochem Biophys* **405**, 130-136 (2002).

24 Tholl, D., Croteau, R. & Gershenzon, J. Partial purification and characterization of the short-chain prenyltransferases, geranyl diphosphate synthase and farnesyl diphosphate synthase, from Abies grandis (grand fir). *Arch Biochem Biophys* **386**, 233-242 (2001).

25 Phillips, M. A., Savage, T. J., & Croteau, R. Monoterpene synthases of loblolly pine (Pinus taeda) produce pinene isomers and enantiomers, *Arch Biochem Biophys* **372**, 197-204 (1999).

26 Lewinsohn, E., Gijzen, M., & Croteau, R. Wound-inducible pinene cyclase from grand fir: purification, characterization, and renaturation after SDS-PAGE, *Arch Biochem Biophys* **293**, 167-173 (1992).

27 Schmidt, A., & Gershenzon, J. Cloning and characterization of two different types of geranyl diphosphate synthases from Norway spruce (Picea abies), *Phytochemistry* **69**, 49-57 (2008).

28 Chang, T. H., et al. Structure of a heterotetrameric geranyl pyrophosphate synthase from mint (Mentha piperita) reveals intersubunit regulation, *Plant Cell* **22**, 454-467 (2010) .

29 Hsieh, F. L., Chang, T. H., Ko, T. P. & Wang, A. H. Structure and mechanism of an Arabidopsis medium/long-chain-length prenyl pyrophosphate synthase. *Plant Physiol* **155**, 1079-1090 (2011).

30 Reiling, K. K., et al. Mono and diterpene production in Escherichia coli, *Biotechnol Bioeng* **87**, 200-212 (2004).

31 Degenhardt, J., Kollner, T. G., & Gershenzon, J. Monoterpene and sesquiterpene synthases and the origin of terpene skeletal diversity in plants, *Phytochemistry* **70**, 1621-1637 (2009).

32 Phillips, M. A., Wildung, M. R., Williams, D. C., Hyatt, D. C., & Croteau, R. cDNA isolation, functional expression, and characterization of (+)-alpha-pinene synthase and (-)-alpha-pinene synthase from loblolly pine (Pinus taeda): stereocontrol in pinene biosynthesis, *Arch Biochem Biophys* **411**, 267-276 (2003).

33 Bohlmann, J., Steele, C. L., & Croteau, R. Monoterpene synthases from grand fir (Abies grandis). cDNA isolation, characterization, and functional expression of myrcene synthase, (-)-(4S)-limonene synthase, and (-)-(1S,5S)-pinene synthase, *J Biol Chem* **272**, 21784-21792 (1997).

34 Martin, D. M., Faldt, J., & Bohlmann, J. Functional characterization of nine Norway Spruce TPS genes and evolution of gymnosperm terpene synthases of the TPS-d subfamily, *Plant Physiol* **135**, 1908-1927 (2004).

35 Kelley, L. A. & Sternberg, M. J. Protein structure prediction on the Web: a case study using the Phyre server. *Nat Protoc* **4**, 363-371 (2009).

36 Hyatt, D. C., et al. Structure of limonene synthase, a simple model for terpenoid cyclase catalysis, *Proc Nat Acad Sci USA* **104**, 5360-5365 (2007) .

37 Ozaydin, B., Burd, H., Lee, T. S., & Keasling, J. D. Carotenoid-based phenotypic screen of the yeast deletion collection reveals new genes with roles in isoprenoid production, *Metab Eng* **15**, 174-183 (2013).

38 Wang, C., et al. Metabolic engineering of Escherichia coli for alpha-farnesene production, *Metab Eng* **13**, 648-655 (2011).

39 Tokuhiro, K., et al. Overproduction of geranylgeraniol by metabolically engineered Saccharomyces cerevisiae, *App Environ Microbiol* **75**, 5536-5543 (2009).

40 McAndrew, R. P., et al. Structure of a three-domain sesquiterpene synthase: a prospective target for advanced biofuels production, *Structure* **19**, 1876-1884 (2011).

41 Williams, D. C., McGarvey, D. J., Katahira, E. J., & Croteau, R. Truncation of limonene synthase preprotein provides a fully active 'pseudomature' form of this monoterpene cyclase and reveals the function of the amino-terminal arginine pair, *Biochemistry* **37**, 12213-12220 (1998).

42 Medicis, E. D., Paquette, J., Gauthier, J. J., & Shapcott, D. Magnesium and manganese content of halophilic bacteria, *Appl Environl Microbiol* **52**, 567-573 (1986).

43 Yoshikuni, Y., Ferrin, T. E., & Keasling, J. D. Designed divergent evolution of enzyme function, *Nature* **440**, 1078-1082 (2006).

44 Hyatt, D. C., & Croteau, R. Mutational analysis of a monoterpene synthase reaction: altered catalysis through directed mutagenesis of (-)-pinene synthase from Abies grandis, *Arch Biochem Biophys* **439**, 222-233 (2005).

45 Lewis, N. E., Nagarajan, H. & Palsson, B. O. Constraining the metabolic genotype-phenotype relationship using a phylogeny of in silico methods. *Nat Rev Microbiol* **10**, 291–305 (2012).

46 Klein-Marcuschamer, D., Oleskowicz-Popiel, P., Simmons, B. A., & Blanch, H. W. Technoeconomic analysis of biofuels: A wiki-based platform for lignocellulosic biorefineries, *Biomass Bioenerg* **34**, 1914-1921 (2010).

47 Appeal of Raython Missile Systems, ASBCA No 5794, 11-2 B.C.A. 3854 (Armed Srv. Bd. of Cont. App. Oct. 11, 2011), available at www.asbca.mil/Decisions/2011/57594_101111_WEB.pdf

48 Sampedro, L., Moreira, X., Llusia, J., Penuelas, J. & Zas, R. Genetics, phosphorus availability, and herbivore-derived induction as sources of phenotypic variation of leaf volatile terpenes in a pine species. *J Exp Bot* **61**, 4437-4447 (2010).

49 Li, M. Z. & Elledge, S. J. Harnessing homologous recombination in vitro to generate recombinant DNA via SLIC. *Nat Methods* **4**, 251-256 (2007).

50 Reed, J. L., Vo, T. D., Schilling, C. H. & Palsson, B. O. An expanded genome-scale model of Escherichia coli K-12 (iJR904 GSM/GPR). *Genome Biol* **4**, R54 (2003).

51 Feist, A. M. et al. A genome-scale metabolic reconstruction for Escherichia coli K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information. *Mol Syst Biol* **3**:121 (2007).

52 Becker, S. A. et al. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox. *Nat Protoc* **2**, 727–738 (2007).

53 Schellenberger, J. et al. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. *Nat Protoc* **6**, 1290–1307 (2011).

CHAPTER 2. MICROBIAL SYNTHESIS OF MEDIUM-CHAIN CHEMICALS FROM RENEWABLES

2.1 Abstract

Linear, medium-chain (C8–C12) hydrocarbons are important components of fuels as well as commodity and specialty chemicals. As industrial microbes do not contain the endogenous pathways to produce medium-chain chemicals, such approaches as overexpression of endogenous enzymes or deletion of competing pathways are not available to the metabolic engineer; instead, fatty acid synthesis and reversed β-oxidation are manipulated to synthesize medium-chain chemical precursors. Even so, chain lengths remain difficult to control, which means that purification must be used to obtain the desired products, titers of which are typically low and rarely exceed milligrams per liter. By engineering the substrate specificity and activity of the pathway enzymes that generate the fatty acyl intermediates and chain-tailoring enzymes, researchers can boost the type and yield of medium-chain chemicals. Development of technologies to both manipulate chain-tailoring enzymes and to assay for products promises to enable the generation of g/L yields of medium-chain chemicals.

2.2 Introduction

Chemicals with chain lengths of eight to twelve carbons are used as biofuels and commodity and fine chemicals (Table 2-1). Medium-chain fatty acids (MCFAs) are used as herbicides, antimicrobials and intermediates for lubricant synthesis, while medium-chain (MC) ketones, such as undecanone, are used as flavors and floral fragrances. MCFA alkyl esters are used in the flavor and fragrance industry, for example ethyl octanoate, which lends aroma to wine[1], while also having a cetane number and cold properties similar to D2 diesel[2]. MC alcohols are used as surfactant additives in diesel blends[3] and in drug delivery. MC alkanes are "drop in" replacements for gasoline and jet fuel[4], while MC α-olefins are used to produce detergents and plasticizers, and as monomers for elastomers used in automotive parts[5]. MC diacids are precursors to polymers[6], such as nylon, and are used in the preparation of perfumes, adhesives, and antibiotics. Finally, MC ω-hydroxyacids serve as monomers for the synthesis of elastomers and bioplastics with similar or improved mechanical and thermal properties over polyethylene[7].

Table 2-1: Medium-chain chemical uses and highest titers. Prices were obtained from Sigma-Aldrich.

	Chain-length	Use	Price ($/g)	Engineered Pathway	Highest Titers	Chassis	Ref
Fatty acids	C8	Herbicide & pesticide	0.06	Reversed β-oxidation	~500 mg/L	*E. coli*	27
	C10	Component in germicidals	0.34	Fatty acid	~500 mg/L	*Y. lipolytica*	37
	C12	Component of topical antimicrobials	1.36	Fatty acid	~467 mg/L	*E. coli*	34
Alkanes	C8	Gasoline alternative	0.5	-	-	-	-
	C9	Gasoline/jet fuel alternative	1.71	Fatty acid	328 mg/L	*E. coli*	4
	C10	Jet fuel alternative	0.42	-	-	-	-

Table 2-1: Continued

	Chain-length	Use	Price ($/g)	Engineered Pathway	Highest Titers	Chassis	Ref
Alkanes	C11	Jet fuel alternative	1	Fatty acid	0.078 mg/L/OD	*E. coli*	42
α-olefins	C8	Monomer for ethylene-α-olefin elastomer	0.48	-	-	-	-
	C11	Monomer in poly-α-olefins for lubricants	8.58	Fatty acid	55 mg/L	*E. coli*	17
Alkyl esters	C8 (ethyl)	Flavor/odor agent (apricot/banana/ floral/pear/)	2.68	Reversed β-oxidation	~0.4 mg/L	*S. cerevisiae*	15
	C10 (ethyl)	Biodiesel	0.33	Fatty acid	22 mg/L	*E. coli*	4
	C12 (ethyl)	Flavor/odor agent (fruity/floral)	0.64	Fatty acid	363 mg/L	*E. coli*	4
	C8 (isobutyl)	Surfactant	51.5	Fatty acid	~6 mg/L	*E. coli*	49
	C10 (isobutyl)	Surfactant	51.5	Fatty acid	~6 mg/L	*E. coli*	49
	C12 (isobutyl)	Component in polyurethane adhesives	51.5	Fatty acid	~35 mg/L	*E. coli*	49
	C10 (isoamyl)	Biodiesel	51.5	Fatty acid	~12.5 mg/L	*S. cerevisiae*	50
	C12 (isoamyl)	Flavor/odor agent (Beer/brandy/ creamy/ grape/)	0.99	Fatty acid	~12.5 mg/L	*S. cerevisiae*	50
	C12 (methyl)	Flavor/odor agent (Coconut/creamy/ fatty/soapy)	10.08	Fatty acid	~250 mg/L	*E. coli*	52
ω-hydroxyacids	C8	Used in synthesis of certain HDAC inhibitors	875	Reversed β-oxidation	403 mg/L	*E. coli*	35
	C10	Precursor to 10-acryloyloxydecanoate, for copolymers	127	Reversed β-oxidation	150 mg/L	*E. coli*	35
	C12	Monomer of poly(12-hydroxydodecanoates) for elastomers	38.9	Fatty acid	111 mg/L	*E. coli*	60
α,ω-hydroxyacids	C8	Precursor to 1,8-octanediamine: used in polyamides	11.4	Reversed β-oxidation	254 mg/L	*E. coli*	35
	C10	Monomer in commercial nylon-6,10	5.22	Reversed β-oxidation	61 mg/L	*E. coli*	35

Table 2-1: Continued

	Chain-length	Use	Price ($/g)	Engineered Pathway	Highest Titers	Chassis	Ref
α,ω-hydroxyacids	C11	Used as a corrosion inhibitor	117	Fatty acid	~5 mg/L	*E. coli*	62
	C12	Monomer in commercial nylon-6,12	0.76	Fatty acid	567 mg/L	*E. coli*	60
Alcohols	C8	Used in diesel blends	0.08	Reversed β-oxidation	~80 mg/L	*E. coli*	24
	C10	Used for drug delivery	0.12	Fatty acid	550 mg/L	*Y. lipolytica*	37
	C11	Active ingredient in some acaricides and nematocides	0.66	Fatty acid	~363 mg/L	*E. coli*	66
	C12	Surfactant to improve the phase stability of diesel blends	1.36	Fatty acid	~1.25 g/L	*E. coli*	64
Methyl ketones	C11	Flavor/odor agent (citrus/fruity/rose)	7.27	Modified β-oxidation	~57 mg/L	*E. coli*	25
	C13	Flavor/odor agent (spicy/herbaceous)	2.51	Modified β-oxidation	~137 mg/L	*E. coli*	25

MC chemicals are overwhelmingly produced from petroleum, so processes to produce them from renewable feedstocks are urgently needed. Glucose, the main component of cellulose, is used for MC chemical production by engineered industrial bacteria and yeast**.** Advances in lignin isolation and depolymerization may enable conversion of lignin monomers to MC chemicals in the near future (Figure 2.1). Alternative microbial chasses, such as *Ralstonia eutropha*[8] and cyanobacteria[9] are also being engineered, with the aim of converting hydrogen, carbon dioxide and sunlight into MCFAs.

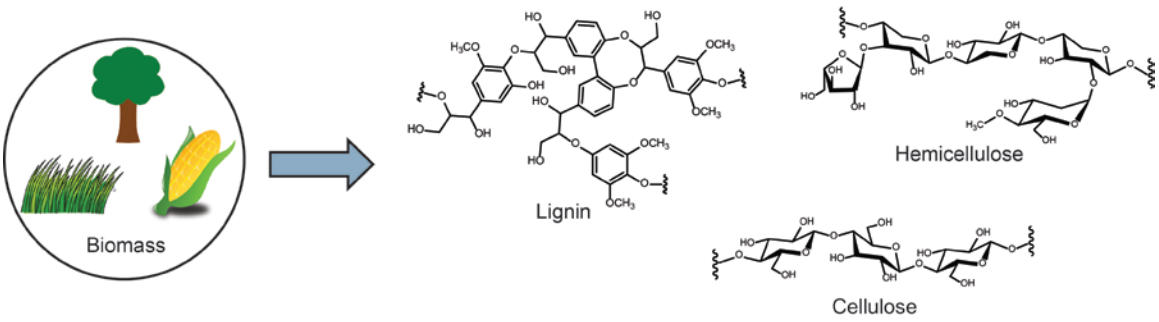

Figure 2.1: Production of chemicals from lignocellulosic biomass. Lignocellulosic biomass comprises the glucose polymer cellulose, hemicellulose and lignin. Depending on the biomass source, cellulose can constitute from 90% (cotton) to 50% (wood) of the biomass. The ease of depolymerization of cellulose by cellulase into glucose monomers, and the fact that it forms the largest portion of lignocellulosic biomass, makes it the main substrate for the microbial production of chemicals. Hemicellulose, a polymer composed of C6 and C5 sugars, is challenging to depolymerize due to the different reactivity of the linkages in the polymer. The main component of hemicellulose is the C5 sugar, which can be naturally used for fermentation and chemical production by *Escherichia coli*[10], but not *Saccharomyces cerevisiae*. About 35% of lignocellulosic biomass is composed of lignin, a heterogenous phenolic polymer, which is increasingly being used to produce chemicals. Lignin monomers, such as coumaric acid and ferulic acid, can be funneled to key metabolic intermediates, such as acetyl-CoA, to produce value added chemicals in *Pseudomonas*[11]. Key challenges in the deconstruction of lignocellulosic biomass to simple sugars include the efficient and cost-effective delignification of biomass, the cost and activity of enzymes to break down cellulose and hemicellulose into sugar monomers, and the discovery and engineering of enzymes able to break down lignin into lignin monomers. Plant biomass is comprised of varying proportions of lignin, glucose and cellulose.

In cells, fatty acid synthases (FASs) produce MC fatty acyls, which are key precursors of MC chemicals. FASs have evolved to cycle until fatty acyls achieve C16–C18 carbons in length before being released. Hence, FAS release only very small amounts of MC fatty acyl intermediates to the cytosol[12]. Finding ways to release fatty acyl intermediates of defined lengths as products is one of the main challenges facing metabolic engineers. Production of MC fatty acyl precursors is only the first step. Precursors are processed by chain-tailoring enzymes, such as wax ester synthases, reductases and monooxygenases, to produce MC chemicals. Chain-tailoring enzymes

often prefer long-chain precursors, and must be engineered to accept MC precursors as well as maintain wild-type enzyme activity, or yields of MC chemicals will be too low.

Despite a handful of reports in which MC chemicals of a single chain length predominate, most studies output a mixture of chain lengths. This reduces the yield of the desired MC chemical, which is compounded by difficulty in purification, because MC chemicals of different lengths have similar physicochemical properties.

In the past two years, enzymes in the fatty acid and the β-oxidation pathways have been engineered to narrow the range of fatty acyl intermediates produced. Chain-tailoring enzymes are also being engineered to further reduce the range of chain lengths produced (Table 2-2). We now need to understand the kinetics and structures of chain-tailoring enzymes to inform protein engineering efforts to alter substrate specificity or improve function. Finally, a lack of high-throughput screens for MCFAs, and, more broadly, MC chemicals, hinders the application of mutagenesis strategies to engineer chain-tailoring enzymes with the properties desired for use in microbial production strains. We review metabolic and protein engineering strategies to produce MC chemicals from renewables, with a focus on approaches to optimize production of specific chain lengths.

Table 2-2: Enzymes engineered for MC chemical production. a. *In vitro*

Chemical	Enzyme	Outcome	Ref
Fatty acids	*E. coli* FAS KS (FabF):I108F	32% improvement in octanoic acid titers	13
	S. elongates ACP:I75Y/I57M	2-fold increase in total free fatty acid titers	21
	C. ammoniagenes FAS KS:G2599S/M2600W; AT:I151A	11-fold increase in C8-acyl coA levels[a]	76

Table 2-2: Continued

Chemical	Enzyme	Outcome	Ref
Fatty acids	*S. cerevisiae* FAS KS:G1250S; MPT:R1834K; AT: I306A	~27-fold increase in C6 and C8 acids over wild type	22
	H. sapiens FAS: thioesterase replaced with *R. norvegicus* thioesterase	~55-fold increase in octanoic acid titers	14
	Y. lipolytica FAS: MPT replaced with *U. californica* thioesterase	6-fold increase in dodecanoic acid	23
	S. cerevisiae FAS KS:G1250S/M1251W & ACP fused to *A. baylyi* thioesterase	~4.2-fold increase in C6-C12 fatty acids	20
	E. coli thioesterase 'tesA:M141L/Y145K/L146K	10-fold improvement in C8 mole fraction over wild type	33
	E. coli tesA':L109P	~12.8-fold increase in C12 fatty acid titers and ~2.2-fold increase in C10 fatty acid titers	4
	E. coli tesA:D17S/L162R	2.2-fold improvement of activity for C12-CoA substrates[a]	77
Alkanes	*P. marinus* ADO:A134F	2-fold improved turnover of C6 and C8 aldehydes[a]	41
	S. elongatus ADO:V184F	Enabled the production of 0.078 mg/L/OD of undecane	42
Esters	*M. aquaeolei* WS:M405F, WS:M405W	~2.2-fold improved selectivity towards isoamyl alcohol and ~2.8-fold towards ethanol[a]	51
ω-hydroxyacids	*M. aquaoelei* CYP153A:G307A	Contributed to the production of 111 mg/L of 12-hydroxydodecanoic acid	60
	M. aquaeolei CYP153A:G307A/S233G	2-fold improved selectivity towards dodecanoic acid and 14-fold for octanoic acid[a]	78
	M. aquaeolei CYP153A:A231G	3-fold improved conversion to12-hydroxydodecanoic acid[a]	79

2.3 Medium-chains in nature

In nature, MCFAs are present in very low concentrations in organisms that aren't amenable for industrialization, such as seeds and the fruit of palm and coconut. Plant MC thioesterases have been expressed in *Escherichia coli* and *Saccharomyces cerevisiae* for MCFAs production[13, 14]. With respect to MCFA esters, *S. cerevisiae* has endogenous

acyl-CoA/ethanol *O*-acyltransferases and produces ethyl octanoate and ethyl decanoate for fatty acid detoxification[1]. However, biotechnological manipulation of acyl-CoA/ethanol *O*-acyltransferases has only produced μg/L titers so far[15]. MC alkanes are naturally produced by *Saccharomyces oviformis* and *Cladosporium resinae*; but pathways that manufacture these alkanes remain obscure[16]. *Pseudomonas* produces the antimicrobial undecene[17, 18], and undecene decarboxylase has been used to produce undecene in *E. coli*[17]. MC hydroxyacids, are synthesized by *Pseudomonas, Marinobacter, Acinetobacter,* and *Mycobacterium* among others[7] and could be a source of useful pathways and genes. MC methyl ketones in tomato are used as insecticides[19], and *S. cerevisiae* was engineered using the tomato thioesterase to synthesize 2-undecanone and 2-tridecanone[20]. Understanding these and other naturally occurring MC chemical biosynthetic pathways will enable the discovery of enzymes to produce high levels of MC chemicals in industrially relevant microbes.

2.4 Engineering medium-chain chemical precursors

Precursors for MC chemicals have been generated using fatty acid, reverse β-oxidation, terpene, and amino acid pathways (Figure 2.2).

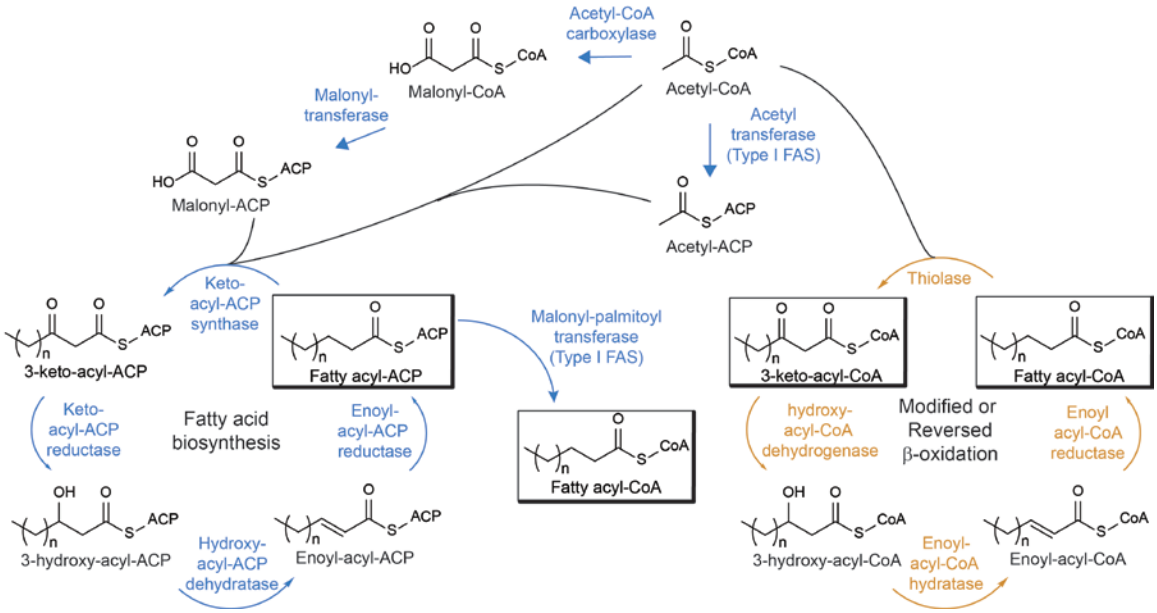

Figure 2.2: Precursor pathways for MC chemicals. Fatty acid biosynthesis (blue) and modified or reversed β-oxidation pathway (orange) produce C8-C12 acyl-CoA/ACP intermediates (n= 4, 6, 8). Precursors are extended two carbon units at a time through condensation of acetyl-CoA (reversed β-oxidation) or malonyl-CoA (fatty acid biosynthesis) followed by reduction, dehydration, and a second reduction. Boxed intermediates are key precursors to MC chemicals.

2.4.1 Fatty acid biosynthesis

Type I FAS (a multienzyme complex found in eukaryotes) or Type II FAS (multiple discrete enzymes found in bacteria) synthesize fatty acyl intermediates. In FAS, the keto synthase (KS), thioesterase, malonyl-palmitoyl transferase (MPT), and acetyl transferase (AT) determine the fatty acyl chain length. Additionally, the acyl-carrier protein (ACP), which carries fatty acyl intermediates from one enzyme active site to another, is also a key chain length determinant. Both, Type I and Type II FAS have been engineered to increase MC fatty acyl yields[13, 21, 22]. In Type II FAS, shrinking the size of the KS (FabF) active site or decreasing the ACP acyl chain binding pocket increases MC fatty acyl production. For example, FabF elongates C6-C14 acyl-ACPs, whereas FabF:I108F

mainly elongates C6 acyl-ACP[13]. Expression of *Synechococcus elongatus* ACP:I75Y/I57M in *E. coli* doubled free fatty acid titers[21]. In Type I FAS, mutating the KS and AT binding pockets enables production of MC fatty acyls. Replacing wild-type *S. cerevisiae* FAS with a mutated version (KS:G1250S; MPT:R1834K;AT:I306A) enabled production of 245 mg/L of octanoic acid (**1**)[22]. Domain swapping can also modulate fatty acyl chain length. Swapping the *Yarrowia lipolytica* FAS MPT with *Umbellularia californica* thioesterase increased dodecanoic acid 6-fold when compared with cytosolic expression *of U. californica* thioesterase[23]. Analogously, swapping the human FAS thioesterase with that of rat and expressing it in *S. cerevisiae* increased octanoic acid 55-fold when compared to wild-type[14]. More recently, localization of *Acinetobacter baylyi* thioesterase inside an *S. cerevisiae* FAS harboring KS:G1250S/M1251W enabled the thioesterase to act directly on fatty acyl-ACPs, resulting in 175 mg/L of C6-C12 fatty acids[20].

2.4.2 Reversed β-oxidation pathway

The β-oxidation pathway degrades fatty acyl-CoAs to acetyl-CoA, and when run in reverse it can synthesize fatty acyl-CoAs. The main challenge is that reversed β-oxidation will only turn for one cycle, which mainly produces C4 fatty acyls, which in turn reduces the yield of MCFAs and MC chemicals, making downstream purification a necessity. Disrupted and reversed β-oxidation have been used to engineer *E. coli*[24, 25] and *S. cerevisiae*[15, 26] for MC chemical production. In reversed β-oxidation, the thiolase determines the fatty acyl-CoA chain length[27] but this enzyme has not yet been engineered for tailored product chain length.

2.4.3 Terpene biosynthesis

The linear monoterpenes (C10) geraniol, linalool and myrcene (2-4) have been produced in bacteria and yeast. Expression of *Valeriana officinalis* geraniol synthase together with dynamic regulation of geraniol diphosphate synthase resulted in 1.69 g/L of geraniol in *S. cerevisiae*[28]. *E. coli* carrying an engineered eukaryotic mevalonate pathway expressing a farnesyl diphosphate synthase mutant capable of producing geraniol diphosphate and *Mentha citrate* linalool synthase produced 505 mg/L of linalool[29]. Myrcene has also been produced in *E. coli* carrying the same eukaryotic mevalonate pathway co-expressing *Abies grandis* geraniol diphosphate synthase and *Quercus ilex* myrcene synthase to yield 58 mg/L of myrcene[30].

2.4.4 Amino acid biosynthesis

Amino acid production has been engineered in bacteria to achieve near theoretical yields and are therefore excellent precursors for MC chemical production [31]. In *E. coli*, octanol (**5**) was produced at 10 mg/L by a synthetic recursive +1C pathway. The pathway starts at threonine, elongates the threonine 2-ketoacid using leucine pathway enzymes, and terminates using a ketoisovalerate decarboxylase and an alcohol dehydrogenase[32].

2.5 From medium-chain precursors to products

Most MC chemicals can be derived from MCFAs (Figure 2.3). Strategies to overproduce MCFAs include altering the substrate specificity of FAS chain-length controlling enzymes, limiting the number of fatty acyl extension cycles and engineering non-FAS pathways, such as reverse β-oxidation. An additional wrinkle is that MCFAs

are toxic to the cell and long-chain fatty acids are essential for cell wall biosynthesis and posttranslational modifications. Therefore, long-chain fatty acyl production cannot be deleted in producer strains. Rather, medium- and long-chain fatty acyl production must be balanced to enable host growth while routing as much carbon as possible into MC chemical production.

Figure 2.3: Metabolic pathways for medium-chain (MC) chemicals. Fatty acid biosynthesis (blue), and modified and reversed β-oxidation pathways (orange) generate fatty acyl-intermediate precursors that are modified by chain-tailoring enzymes (green) into MC chemicals. Other MC chemical producing pathways are amino acid and terpene biosynthesis (purple). n= 5-9 for all MC chemicals except for alkanes, α-olefins, and α,ω-diacids where n = 4-8.

2.5.1 Fatty acids

Fatty acids are produced via hydrolysis of fatty acyl acyl-ACPs or acyl-CoAs by thioesterases. Bacterial acyl-ACP thioesterases have been engineered for narrower chain lengths. The *E. coli* thioesterase TesA hydrolyzes C14:0 and C16:0 acyl-ACPs, while

TesA:L109P hydrolyzes C12:0[4]. Recently, TesA:M141L/Y145K/L146K showed a 10-fold improvement in C8 mol fraction over wild type[33]. The most successful MCFA productions in *E. coli* have leveraged disruption[13, 34] or reversal[27, 35] of β-oxidation coupled to overexpression of MC thioesterases. Octanoic acid has been produced at ~500 mg/L by using reversed β-oxidation and overexpressing TesA[27]. Decanoic acid (**6**) has been produced at 250 mg/L by independently controlling the expression of thiolase, ketosynthase and trans-enoyl-CoA reductase in reversed β-oxidation together with expression of *E. coli* thioesterase ydiI[35]. Dodecanoic acid (**7**) has been produced at ~467 mg/L via disruption of β-oxidation, overexpression of *E. coli* acetyl-CoA carboxylase (ACC) to increase flux via fatty acid biosynthesis, and expression of the MC *U. californica* thioesterase[34]. More recently, fine-tuning the expression of acetyl-CoA synthetase to convert acetate back into acetyl-CoA and the thiolase in conjunction with repression of byproduct pathways via CRISPRi resulted in 3.8 g/L of C6–C10 acids[36].

Eukaryotes have also been engineered for MCFA production. Although Type I FAS is more kinetically efficient than Type II FAS due to the proximity of the catalytic domains in the complex and the high concentration of fatty acyl-ACP in the chamber, the complex physically blocks access by MC thioesterases to fatty acyl-ACPs. Consequently, in eukaryotes, the most successful strategies have been mutations to FAS chain-length determining enzymes, and expression of MC thioesterases fused to FAS. In *S. cerevisiae*, octanoic acid has been produced at 245 mg/L by mutations in KS, MPT, and AT and relying on endogenous thioesterases[22]. Decanoic acid and dodecanoic acid have been produced at ~50 mg/L and ~10 mg/L, respectively, by localizing the MC *A. baylyi* thioesterase inside the FAS chamber, via fusion to the *S. cerevisiae* ACP domain,

enabling the thioesterase to directly act on fatty acyl-ACPs, while mutating KS to prevent elongation of MC acyl-ACPs[20]. Work the oleaginous yeast *Y. lipolytica* suggest that small thioesterases may diffuse into the *Y. lipolytica* FAS chamber and directly act on fatty acyl-ACP pools[23]. Expression of *Cuphea palustris* acyl-ACP thioesterase in *Y. lipolytica* resulted in ~500 mg/L of decanoic acid[37].

Non-traditional chassis lines, such as cyanobacteria and *Ralstonia,* have the potential to convert CO_2 or hydrogen to MCFAs and have been engineered to convert sugars to MCFAs. In cyanobacteria, overexpression of *Cuphea hookeriana* thioesterase and the endogenous ACC coupled to disruption of cyanophycin synthesis resulted in ~2.1 mg/L of decanoic acid and ~19 mg/L of dodecanoic acid[9]. In *R. eutropha*, overexpression of the *U. californica* thioesterase results in 14 mg/L of dodecanoic acid[8].

MCFAs are toxic to bacteria and yeast because they damage membranes resulting in metabolite leakage and generating reactive oxygen species[38, 39]. In *E. coli*, addition of ~5 g/L of MCFAs results in complete growth inhibition[38]; so at current MCFA titers of ~500 mg/L in this chassis, chemical toxicity is not a limiting factor. In *S. cerevisiae* however addition of ~100mg/L of MCFAs results in complete growth inhibition[39]. *S. cerevisiae* MCFAs titers are currently >100 mg/L, therefore chemical toxicity may limit MCFA production in this chassis.

2.5.2 Alkanes

Deformylation of fatty aldehydes by aldehyde deformylating oxygenases (ADO) results in odd chain alkanes. Nonane (**8**) has been produced to 328 mg/L in *E. coli* expressing TesA:L109P, *Clostridium acetobutylicum* acyl-CoA reductase, and *A.*

thaliana ADO[4]. The main obstacle in alkane production is ADO's slow turnover rate (~7 h^{-1})[40] and broad substrate specificity, deformylating both medium- and long- chain aldehydes[4]. *In vitro*, *P. marinus* ADO:A134F has a 2-fold improved turnover of C6 and C8 aldehydes, while being inactive with long-chain aldehyde[41]. Also *in vitro*, *S. elongatus* ADO:V184F has a 4.4-fold improved undecane (**9**) production over wild-type. *S. elongatus* ADO:V184F has been used in *E. coli* to produce 0.078 mg/L/OD of undecane[42]. MC alkanes have been recently produced in *S. cerevisiae* using engineered FAS KS:G1250S/M1251W and expression of *Thermosynechococcus elongatus* BP-1 ADO generated ~30 μg/L/OD of nonane and ~5 μg/L/OD of undecane[43].

2.5.3 α-olefins

Decarboxylation of fatty acids by desaturase-like enzymes, which prefer C10–C16 acids[17], fatty acid decarboxylases, which prefer C4–C18 acids[44], or iron oxidases, which prefer C6–C10 acids[17] results in α-olefins. Expression of a desaturase-like enzyme in *E. coli* resulted in 55 mg/L of undecene (**10**)[17], while expression of a fatty acid decarboxylase in *S. cerevisiae* resulted in ~22 μg/L of undecene[45]. Use of iron oxidases has resulted in ~5 mg/L of MC α-olefins in *E. coli*[17]. Only the structure of fatty acid decarboxylase[46] is known. Elucidating the structure of iron oxidase and desaturase-like enzymes and identifying more efficient reductases should aid in increasing MC α-olefins titers.

2.5.4 Alkyl esters

Coupling of fatty acyl-CoAs and alcohols by wax ester synthase (WS) result in alkyl esters. WS naturally couples long-chain fatty acyl-CoAs and medium- to long-chain

alcohols[47], and WSs need to be engineered to more efficiently couple MC fatty acyl-CoAs and short chain alcohols for MC alkyl ester production. In an ethanol-producing *E. coli*, expression of *Acinetobacter* sp. ADP1 WS produced ethyl decanoate (**11**) and ethyl dodecanoate (**12**) at 22 mg/L and 363 mg/L, respectively[4]. In *S. cerevisiae*, a natural ethanol producer, the same molecules were produced at 15 mg/L and 5 mg/L, respectively, using *Marinobacter hydrocarbonoclasticus* WS[48]. In *Y. lipolytica*, overexpression of *A. baylyi* WS and a peroxisomal/mitochondrial carnitine acyltransferase to shuttle acetyl-CoA from the main reservoirs in the peroxisome and mitochondria to the cytosol resulted in ~12.8 mg/L of ethyl dodecanoate[23]. Coupling fatty acyl-CoAs and branched alcohols results in branched alkyl esters. An isoamyl/isobutanol-producing *E. coli* expressing *Acinetobacter sp.* WS produced isobutyl dodecanoate (**13**), isobutyl octanoate (**14**), and isobutyl decanoate (**15**) at ~47 mg/L[49]. An isobutanol-producing *S. cerevisiae* overexpressing *Marinobacter sp.* WS produced isobutyl/isoamyl decanoate (**16**) and isobutyl/isoamyl dodecanoate (**17**) at ~25 mg/L[50]. WS engineering is limited by the lack of crystal structure and low homology to available structures. After using bioinformatics to identify the WS active site, *M. aquaeolei* WS:M405W and WS:M405F was engineered to have a ~2.8-fold improved selectivity towards ethanol and ~2.2-fold improved selectivity for isoamyl alcohol, respectively, when compared to wild type[51]

Methyl esters are generated via fatty acid *O*-methyl transferases (OMT) using *S*-adenosyl methionine as the methyl group donor. Expression of *Drosophila melanogaster* OMT in *E. coli* resulted in ~250 mg/L of methyl dodecanoate (**18**)[52]. *D. melanogaster* OMT's endogenous substrate is C15 terpenes, and had 100-fold lower turnover with fatty

acids[53]; thus the challenge is engineering its substrate specificity. Although *M. marinum* OMT prefers decanoic acid, its biotechnology use resulted in only nM levels of decanoic acid[54]. The recently solved *M. marinum* OMT structure[55] should help explain the drivers for its low activity and provide avenues for its engineering.

2.5.5 ω-hydroxyacids

Terminal hydroxylation of fatty acids by cytochrome P450s, such as CYP153A, which prefers C5–C16 acids[7] and CYP52, which prefers C12–C18[56], and non-CYP450 alkane monooxygenases[35] results in ω-hydroxyacids. Use of *P. putida* non-CYP450 alkane monooxygenase in *E. coli* resulted in 275 mg/L of 6-hydroxyhexanoic acid, 403 mg/L of 8-hydroxyoctanoic acid (**19**), and 150 mg/L of 10-hydroxydecanoic acid (**20**)[35]. In a different approach, condensing acetyl-CoA and glucaryl-CoA using reversed phenylacetate degradation coupled to expression of *Clostridium beijerinckii* acyl-CoA reductase and *E. coli* native alcohol dehydrogenases in a strain with deficient mixed acid fermentation resulted in 87 mg/L of 7-hydroxyheptanoic acid (**21**), in a 30% conversion from glucaric acid[57]. In *S. cerevisiae*, *Fusarium oxysporum* CYP539A7 (~40% homology to CYP52) co-expressed with its homologous reductase had a maximum of 67% conversion of C8/C10/C12 acids to 8-hydroxyoctanoic acid, 10-hydroxydecanoic, and 12-hydroxydodecanoic acid (**22**)[58]. In *Y. lipolytica*, an endogenous CYP52 had a 2.4% conversion of dodecane to 12-hydroxydodecanoic acid[59]. CYP153A has been engineered toward shorter chain substrates. Expression of CYP153A:G307A in *E. coli* resulted in 111 mg/L of 12-hydroxydodecanoic acid[60]. CYP153As are likely the best engineering target due to their regioselectivity for the ω-position on fatty acids, unlike CYP52s, and

their tendency to not further oxidize ω-hydroxyacids to other chemicals as it is the case with CYP52s and non-P450 monooxygenases[61].

2.5.6 α, ω-diacids

Oxidation of ω-hydroxyacids to ω-oxoacids by alcohol dehydrogenases (ADHs) followed by a second oxidation of the ω-oxoacids by aldehyde dehydrogenase (ALDHs) results in α,ω-diacids. Bacterial ADHs and ALDHs with broad substrate specificity have been used to produce C6-C12 α,ω-diacids[35, 60]. Suberic acid (C8, **23**) and sebacic acid (C10, **24**) have been produced at 254 mg/L and 61 mg/L, respectively, in *E. coli* using *Acinetobacter* sp. ADH and ALDH[35]. C12 α,ω-diacid has been produced at 567 mg/L by taking advantage of *P. putida* ADH, and a native *E. coli* ALDH[60]. Production of odd chain α,ω-diacids requires a different strategy. Pimelic acid (C7, **25**) was produced at 25 mg/L by coupling acetyl-CoA and glucaryl-CoA using reversed phenylacetate degradation pathway and expressing a thioesterase in a strain with deficient mixed acid fermentation[57]. C11 α,ω-diacid (**26**) has been produced at ~5 mg/L in *E. coli* by shunting biotin synthesis towards diacid production via overexpression of *Kurthia O*-methyltansferase[62]. It remains a challenge to engineer ADH and ALDH for enhanced activity toward MC ω-hydroxy fatty acids.

2.5.7 Alcohols

Fatty acyl-CoAs or fatty acyl-ACPs are sequentially reduced by fatty acyl reductases and aldehyde reductases to alcohols. Dual function reductases can perform this double reduction[63]. Alternatively, fatty acids can be sequentially reduced by carboxylic acid reductases and aldehyde reductases to alcohols. Dual function reductases have been most

successful in the production of MC alcohols[37, 64, 65]. In *E. coli*, decanol (**27**) and octanol have been produced at 170 mg/L and 80 mg/L, respectively, by overexpressing dual function *E. coli* acyl-CoA reductase[24]. Dodecanol (**28**) has been produced at ~1.25 g/L in *E. coli* by using the *M. aquaeolei* acyl-CoA reductase[64]. Undecanol (**29**) has been produced at ~363 mg/L in *E. coli* by overexpressing *Oryza sativa* α-dioxygenase, which turns fatty acids into odd-chain aldehydes, and *E. coli* aldehyde reductase[66]. In *S. cerevisiae*, use of *Tyto alba* acyl-CoA reductase resulted in ~90 mg/L of decanol and ~358 mg/L of dodecanol[65]. In *Y. lipolytica*, the *A. thaliana* acyl-CoA reductase led to 550 mg/L of decanol[37]. Acyl-CoA reductases tend to form inclusion bodies[67] and have broad substrate specificity[68]. Recently, *M. aquaeolei* aldehyde reductase has been crystallized[68], which will aid in the engineering of reductases with improve specificity and expression.

MC alcohols are toxic to bacteria and yeast because they embed themselves in the membrane[69]. Addition of ~550mg/L of MC alcohols to *E. coli* results in growth inhibition[70 69]; however, at higher concentrations MC alcohols form a separate layer. Given that octanol production in *E. coli* is ~100 mg/L, and dodecanol titers have reached > 1g/L, toxicity is likely limiting octanol but not dodecanol production. Octanol and decanol reduce *S. cerevisiae* cell growth at ~500mg/L[71]. As *S. cerevisiae* production of decanol is 90 mg/L, toxicity is likely not limiting alcohol titers.

2.5.8 Methyl ketones

Spontaneous decarboxylation of 3-keto-carboxylic acid, generated by β-ketoacyl-CoA thioesterase in β-oxidation, results in methyl ketones. In *E. coli*, undecanone (**30**) and tridecanone (**31**) have been produced at ~57 mg/L and ~137 mg/L, respectively, using

Micrococcus luteus acyl-CoA oxidase and deleting *fadA* together with expression of native FadB, and thioesterase FadM[25]. In *S. cerevisiae*, ~0.5 μg/g/Dry Cell Weight (DCW) of undecanone and ~1.5 μg/g/DCW of tridecanone have been produced by locating *Solanum habrochaites* β-ketoacyl-ACP thioesterase (ShMKS2) inside the *S. cerevisiae* FAS chamber to act on fatty acyl-ACPs[20]. To date, there are no reports of the engineering of β-ketoacyl-CoA thioesterases. FadM generated ~4-fold more tridecanone than ShMKS2 in the same *E. coli* background[25] FadM has broad specificity for β-ketoacyl-CoA substrates (C10-C18)[25] and displays ~10-fold more activity with saturated C16 acyl-CoA vs C16 β-ketoacyl-CoAs.[72] The availability of FadM crystal structure (PDB:1NJK) will be useful for refining specificity towards MC methyl ketone production.

2.6 Outlook

Microbial production of MC chemicals is only in the mg/L range (Table 2-3, APPENDIX E. theoretical yield calculations). The main challenge is that MC chemicals are not endogenous to biological systems; rather, evolution has optimized the production of long-chain or short-chain fatty acyl intermediates. Although chain-controlling enzymes in FAS are being engineered for narrower substrate/product profile, the main chain-determining enzyme in reversed β-oxidation, thiolase, has not yet been engineered to control chain length. Access to a fatty acyl intermediate pool with a narrower MC-length distribution would help to synthesize single MC chemicals; we foresee substantial efforts aimed at engineering thiolases in future. Synthetic biology tools, such as promoters activated at different stages of the cultivation and dynamic

sensor-regulator systems[73] could help to balance synthesis of MC chemicals with synthesis of essential long chain fatty acids.

Table 2-3: Theoretical Yield (TY), highest reported yields and titers of microbially produced MC chemicals. a. Molar yield from glucose calculated using COBRA toolbox for Matlab. b. Calculated as (moles of product)/(moles of carbon source) c. Theoretical stoichiometric yield is based on carbon chain length of MCFA precursor, does not use carbon from glucose for ethyl side chain. d. Theoretical pathway, involves undiscovered enzymes.

Chemical	Chain length	TY reversed β-oxidation (%)[a]	TY fatty acid biosynthesis (%)[a]	TY Stoichiometric (%)	Reported Yield (%)[b]	Reported titers (mg/L)	Chassis	Reference
Fatty acids	C8	54	51	75	2.6	~500	*E. coli*	27
	C10		43	60	1.1	~500	*Y. lypolytica*	37
	C12	35	33	50	4.2	~467	*E. coli*	34
Alkanes	C9	40	37	60	2.5	328	*E. coli*	4
	C11	33	31	50	-	0.078 mg/L/OD	*E. coli*	42
α-olefins	C11	34	32	50	0.7	55	*E. coli*	17
Ethyl esters	C8	44		75[c]	<<1	0.4	*S. cerevisiae*	15
	C10	36	34	60[c]	0.1	22	*E. coli*	4
	C12	31	29	50[c]	1.6	363	*E. coli*	4
ω-hydroxy acids	C8	52	50	75	1.4	403	*E. coli*	35
	C10	41	39	60	0.4	150	*E. coli*	35
	C12	34	32	50	0.9	111	*E. coli*	60
α,ω-diacids	C8	56	52[d]	75	0.8	254	*E. coli*	35
	C9		37	67	<<1	0.5	*E. coli*	62
	C10	44	41[d]	60	0.2	61	*E. coli*	35
	C11		33	55	<<1	~5	*E. coli*	62
	C12	36	34	50	1.3	~567	*E. coli*	60
Alcohols	C8	50	46	75	3.5	~80	*E. coli*	24
	C10		41	60	1.3	550	*Y. lypolytica*	37
	C11	34	32	50	1.3	363	*E. coli*	66
	C12	33	31	50	9.3	1250	*E. coli*	64
Methyl ketones	C11	37	34	50	0.9	~57	*E. coli*	25
	C13	31	29	43	4.8	~137	*E. coli*	25

Reversed β-oxidation pathway results in the highest maximum theoretical yields for MC chemicals, yet the highest MC chemical experimental titers have been achieved through both fatty acid biosynthesis and reversed β-oxidation. The main disadvantage of

fatty acid biosynthesis is the consumption of ATP during carboxylation of acetyl-CoA to malonyl-CoA. MC chemicals reach higher theoretical yields than their long chain counterparts because fewer cycles of fatty acid synthesis and reversed β-oxidation are required for MC chemical production. The yields for MC chemicals range from ~32–56%, with methyl ketones and α-olefins having the lowest. To date, experimental yields of MC chemicals do not surpass 4% of theoretical yield, while the best titers are in the high mg/L range, leaving much room for improvement.

For most MC chemicals, the roadblock to increased yield is not toxicity. Either MC compounds separate from the growth medium, or titers are not high enough to be toxic. Engineering efflux pumps to secrete potentially toxic MC products to the media is not a pressing need. However, the synthesis of alkanes, α-olefins, fatty methyl esters, α,ω-diacids, and alcohols rely on precursors that diffuse out of the cell, such as fatty aldehydes, fatty acids and ω-hydroxyacids. Thus, engineering transporters for uptake of these precursors from the growth medium could massively increase yields. As yields increase, however, toxicity of some MC chemicals, in particular MCFAs, will become problematic. Engineering chain tailoring enzymes to more efficiently convert MCFAs into desired products will be needed to overcome toxicity. Toxic MC alcohols could be expelled from cells using engineered efflux pumps or by *in situ* removal of the chemical using membranes or resins. All other MC chemicals have higher hydrophobicity and low toxicity due to their ability to form a separate layer.

E. coli has so far proven to be the best chassis for production of most MC chemicals, including ~1.25 g/L of dodecanol. However, the highest decanoic acid and decanol titers (~500 mg/L) with minimal engineering have been achieved in *Y. lipolytica*, making it a

potentially better chassis to produce MC chemicals. Introducing the most successful *S. cerevisiae* FAS mutations into the *Y. lipolytica* FAS could further increase titers. Organisms with accelerated doubling times, such as *Vibrio natriegens*, which doubles every 10 minutes, might enable even higher yields. Other microbial chasses engineered for MC chemical production remain at the proof-of-principle stage, including cyanobacteria and *Ralstonia*; nevertheless, these chasses have the potential to use H_2, CO_2, and light to produce MC chemicals. Archaeal species, which have an isoprenoid-based membrane, might be better suited than *E. coli* for the production of isoprenoid-based MC chemicals.

Activity of the chain-tailoring enzymes is still a problem. So far, only thioesterases, WSs, ADOs, and P450s have been engineered for improved activity or chain length specificity. The available structures of fatty acid decarboxylase[46], acyl-CoA reductase[68], β-ketoacyl thioesterase, and a *O*-methyl transferase[55] will enable the identification of chain length determining residues. The lack of structures for other chain-tailoring enzymes important to α-olefins (iron oxidase, desaturase like enzyme), ω-hydroxyacids (alkane monooxygenase), α,ω-diacids (alcohol dehydrogenase, aldehyde dehydrogenase), and ketones (acyl-CoA oxidase) limits our mechanistic understanding of these reactions, hindering their engineering. Most of these enzymes should not be difficult to purify or crystalize as they are not transmembrane proteins. Further, for some enzymes, e.g. thioesterases, the roadblock is not the availability of structures, but the high price or even lack of commercial substrates (fatty acyl-CoAs, fatty acyl-ACPs) to enable kinetic characterization and the rapid comparison of enzymes from different species. Computational modeling breakthroughs have the potential to generate sufficiently

accurate enzyme models, hopefully bypassing the need for structures and accelerating protein engineering endeavors.

Without structures, engineering of chain-tailoring enzymes will require screening libraries of mutants to identify enzyme variants with improved chain length specificity and activity. MC chemicals lack chromophores and must be analyzed using chromatography-based methods, which have throughput limitations, and hover at 100 samples/day. High-throughput screens (>10^6 samples/day) for MC chemicals, such as G-protein coupled receptor-based MCFA sensors[74], or the transcription factor-based MC alkane sensor[75], could be applied to engineer chain-tailoring enzymes and more rapidly improve MC chemical yields.

Elucidating the structure of the remaining chain-tailoring enzymes to better understand their kinetics and developing rapid assays for MC chemical detection will enable us to achieve much higher yields. Indeed, we anticipate that yields of MC chemicals will increase from mg/L to g/L quantities in the next 5–10 years, making the goal of renewable sources of this important group of products a reality.

2.7 References

1 Saerens, S.M. et al. Parameters affecting ethyl ester production by Saccharomyces cerevisiae during fermentation. *Appl Environ Microbiol* **74**, 454-61 (2008).

2 Knothe, G. "Designer" biodiesel: Optimizing fatty ester (composition to improve fuel properties. *Energy & Fuels* **22**, 1358-1364 (2008).

3 Rajesh Kumar, B. & Saravanan, S. Use of higher alcohol biofuels in diesel engines: A review. *Renewable and Sustainable Energy Reviews* **60**, 84-115 (2016).

4 Choi, Y.J. & Lee, S.Y. Microbial production of short-chain alkanes. *Nature* **502**, 571-574 (2013).

5 Chang, Y.W., Lee, D. & Bae, S.Y. Preparation of polyethylene-octene elastomer/clay nanocomposite and microcellular foam processed in supercritical carbon dioxide. *Polymer International* **55**, 184-189 (2006).

6 Cabrales, L., Calderon, K., Hinojosa, I., Valencia, F. & Abidi, N. Synthesis and characterization of polyesters derived from sebacic acid, hexanediol, and hydroquinone. *International Journal of Polymer Analysis and Characterization* **21**, 718-727 (2016).

7 Malca, S.H. et al. Bacterial CYP153A monooxygenases for the synthesis of omega-hydroxylated fatty acids. *Chem Commun (Camb)* **48**, 5115-5117 (2012).

8 Chen, J.S. et al. Production of fatty acids in Ralstonia eutropha H16 by engineering beta-oxidation and carbon storage. *PeerJ* **3**, e1468 (2015).

9 Liu, X., Sheng, J. & Curtiss, R., 3rd. Fatty acid production in genetically modified cyanobacteria. *Proc Natl Acad Sci U S A* **108**, 6899-904 (2011).

10 Bokinsky, G. et al. Synthesis of three advanced biofuels from ionic liquid-pretreated switchgrass using engineered Escherichia coli. *Proc Natl Acad Sci U S A* **108**, 19949-54 (2011).

11 Abdelaziz, O.Y. et al. Biological valorization of low molecular weight lignin. *Biotechnology Advances* **34**, 1318-1346 (2016).

12 Rock, C.O. & Jackowski, S. Regulation of phospholipid synthesis in Escherichia coli. Composition of the acyl-acyl carrier protein pool in vivo. *J Biol Chem* **257**, 10759-65 (1982).

13 Torella, J.P. et al. Tailored fatty acid synthesis via dynamic control of fatty acid elongation. *Proc Natl Acad Sci U S A* **110**, 11290-5 (2013).

14 Leber, C. & Da Silva, N.A. Engineering of Saccharomyces cerevisiae for the synthesis of short chain fatty acids. *Biotechnol Bioeng* **111**, 347-58 (2014).

15 Lian, J. & Zhao, H. Reversal of the beta-oxidation cycle in Saccharomyces cerevisiae for production of fuels and chemicals. *ACS Synth Biol* **4**, 332-41 (2015).

16 Spakowicz, D.J. & Strobel, S.A. Biosynthesis of hydrocarbons and volatile organic compounds by fungi: bioengineering potential. *Applied Microbiology and Biotechnology* **99**, 4943-4951 (2015).

17 Rui, Z., Harris, N.C., Zhu, X.J., Huang, W. & Zhang, W.J. Discovery of a Family of Desaturase-Like Enzymes for 1-Alkene Biosynthesis. *Acs Catalysis* **5**, 7091-7094 (2015).

18 De Vrieze, M. et al. Volatile Organic Compounds from Native Potato-associated Pseudomonas as Potential Anti-oomycete Agents. *Front Microbiol* **6** (2015).

19 Antonious, G.F., Dahlman, D.L. & Hawkins, L.M. Insecticidal and acaricidal performance of methyl ketones in wild tomato leaves. *Bull Environ Contam Toxicol* **71**, 400-7 (2003).

20 Zhu, Z. et al. Expanding the product portfolio of fungal type I fatty acid synthases. *Nat Chem Biol* (2017).

21 Liu, X., Hicks, W.M., Silver, P.A. & Way, J.C. Engineering acyl carrier protein to enhance production of shortened fatty acids. *Biotechnol Biofuels* **9**, 24 (2016).

22 Gajewski, J., Pavlovic, R., Fischer, M., Boles, E. & Grininger, M. Engineering fungal de novo fatty acid synthesis for short chain fatty acid production. *Nat Commun* **8**, 14650 (2017).

23 Xu, P., Qiao, K.J., Ahn, W.S. & Stephanopoulos, G. Engineering Yarrowia lipolytica as a platform for synthesis of drop-in transportation fuels and oleochemicals. *Proc Natl Acad Sci U S A* **113**, 10848-10853 (2016).

24 Dellomonaco, C., Clomburg, J.M., Miller, E.N. & Gonzalez, R. Engineered reversal of the beta-oxidation cycle for the synthesis of fuels and chemicals. *Nature* **476**, 355-U131 (2011).

25 Goh, E.B., Baidoo, E.E.K., Keasling, J.D. & Beller, H.R. Engineering of Bacterial Methyl Ketone Synthesis for Biofuels. *Appl Environ Microbiol* **78**, 70-80 (2012).

26 Zhou, Y.J. et al. Production of fatty acid-derived oleochemicals and biofuels by synthetic yeast cell factories. *Nat Commun* **7**, 11709 (2016).

27 Kim, S., Clomburg, J.M. & Gonzalez, R. Synthesis of medium-chain length (C6-C10) fuels and chemicals via beta-oxidation reversal in Escherichia coli. *J Ind Microbiol Biotechnol* **42**, 465-475 (2015).

28 Zhao, J.Z. et al. Dynamic control of ERG20 expression combined with minimized endogenous downstream metabolism contributes to the improvement of geraniol production in Saccharomyces cerevisiae. *Microb Cell Fact* **16** (2017).

29 Mendez-Perez, D. et al. Production of jet fuel precursor monoterpenoids from engineered Escherichia coli. *Biotechnol Bioeng* (2017).

30 Kim, E.M., Eom, J.H., Um, Y., Kim, Y. & Woo, H.M. Microbial Synthesis of Myrcene by Metabolically Engineered Escherichia coil. *Journal of Agricultural and Food Chemistry* **63**, 4606-4612 (2015).

31 Wendisch, V.F. Microbial production of amino acids and derived chemicals: synthetic biology approaches to strain development. *Curr Opin Biotechnol* **30**, 51-8 (2014).

32 Mak, W.S. et al. Integrative genomic mining for enzyme function to enable engineering of a non-natural biosynthetic pathway. *Nat Commun* **6** (2015).

33 Grisewood, M.J. et al. Computational Redesign of Acyl-ACP Thioesterase with Improved Selectivity toward Medium-Chain-Length Fatty Acids. *Acs Catalysis* **7**, 3837-3849 (2017).

34 Lennen, R.M., Braden, D.J., West, R.A., Dumesic, J.A. & Pfleger, B.F. A process for microbial hydrocarbon synthesis: Overproduction of fatty acids in Escherichia coli and catalytic conversion to alkanes. *Biotechnol Bioeng* **106**, 193-202 (2010).

35 Clomburg, J.M. et al. Integrated engineering of beta-oxidation reversal and omega-oxidation pathways for the synthesis of medium chain omega-functionalized carboxylic acids. *Metab Eng* **28**, 202-212 (2015).

36 Wu, J.J., Zhang, X., Xia, X.D. & Dong, M.S. A systematic optimization of medium chain fatty acid biosynthesis via the reverse beta-oxidation cycle in Escherichia coli. *Metab Eng* **41**, 115-124 (2017).

37 Rutter, C.D. & Rao, C.V. Production of 1-decanol by metabolically engineered Yarrowia lipolytica. *Metab Eng* **38**, 139-147 (2016).

38 Royce, L.A., Liu, P., Stebbins, M.J., Hanson, B.C. & Jarboe, L.R. The damaging effects of short chain fatty acids on Escherichia coli membranes. *Appl Microbiol Biotechnol* **97**, 8317-27 (2013).

39 Borrull, A., López-Martínez, G., Poblet, M., Cordero-Otero, R. & Rozès, N. New insights into the toxicity mechanism of octanoic and decanoic acids on Saccharomyces cerevisiae. *Yeast* **32**, 451-460 (2015).

40 Eser, B.E., Das, D., Han, J., Jones, P.R. & Marsh, E.N. Oxygen-independent alkane formation by non-heme iron-dependent cyanobacterial aldehyde decarbonylase: investigation of kinetics and requirement for an external electron donor. *Biochemistry* **50**, 10743-50 (2011).

41 Khara, B. et al. Production of propane and other short-chain alkanes by structure-based engineering of ligand specificity in aldehyde-deformylating oxygenase. *Chembiochem* **14**, 1204-8 (2013).

42 Bao, L., Li, J.J., Jia, C., Li, M. & Lu, X. Structure-oriented substrate specificity engineering of aldehyde-deformylating oxygenase towards aldehydes carbon chain length. *Biotechnol Biofuels* **9**, 185 (2016).

43 Zhu, Z. et al. Enabling the synthesis of medium chain alkanes and 1-alkenes in yeast. *Metab Eng* (2017).

44 Dennig, A. et al. Oxidative Decarboxylation of Short-Chain Fatty Acids to 1-Alkenes. *Angewandte Chemie-International Edition* **54**, 8819-8822 (2015).

45 Chen, B., Lee, D.Y. & Chang, M.W. Combinatorial metabolic engineering of Saccharomyces cerevisiae for terminal alkene production. *Metab Eng* **31**, 53-61 (2015).

46 Belcher, J. et al. Structure and biochemical properties of the alkene producing cytochrome P450 OleTJE (CYP152L1) from the Jeotgalicoccus sp. 8456 bacterium. *J Biol Chem* **289**, 6535-50 (2014).

47 Annika Rottig, S.W., Alexander Steinbüchel. In vitro characterization of five bacterial WS/DGAT acyltransferases regarding the synthesis of biotechnologically relevant short-chain-length esters. *European Journal of Lipid Science and Technology* **118**, 124-132 (2015).

48 Thompson, R.A. & Trinh, C.T. Enhancing Fatty Acid Ethyl Ester Production in Saccharomyces cerevisiae Through Metabolic Engineering and Medium Optimization. *Biotechnol Bioeng* **111**, 2200-2208 (2014).

49 Tao, H., Guo, D., Zhang, Y., Deng, Z. & Liu, T. Metabolic engineering of microbes for branched-chain biodiesel production with low-temperature property. *Biotechnol Biofuels* **8**, 92 (2015).

50 Teo, W.S., Ling, H., Yu, A.Q. & Chang, M.W. Metabolic engineering of Saccharomyces cerevisiae for production of fatty acid short- and branched-chain alkyl esters biodiesel. *Biotechnol Biofuels* **8**, 177 (2015).

51 Barney, B.M., Ohlert, J.M., Timler, J.G. & Lijewski, A.M. Altering small and medium alcohol selectivity in the wax ester synthase. *Appl Microbiol Biotechnol* **99**, 9675-84 (2015).

52 Sherkhanov, S., Korman, T.P., Clarke, S.G. & Bowie, J.U. Production of FAME biodiesel in E. coli by direct methylation with an insect enzyme. *Sci Rep* **6**, 24239 (2016).

53 Niwa, R. et al. Juvenile hormone acid O-methyltransferase in Drosophila melanogaster. *Insect Biochem Mol Biol* **38**, 714-20 (2008).

54 Nawabi, P., Bauer, S., Kyrpides, N. & Lykidis, A. Engineering Escherichia coli for biodiesel production utilizing a bacterial fatty acid methyltransferase. *Appl Environ Microbiol* **77**, 8052-61 (2011).

55 Petronikolou, N. & Nair, S.K. Biochemical Studies of Mycobacterial Fatty Acid Methyltransferase: A Catalyst for the Enzymatic Production of Biodiesel. *Chem Biol* **22**, 1480-90 (2015).

56 Huang, F.C., Peter, A. & Schwab, W. Expression and Characterization of CYP52 Genes Involved in the Biosynthesis of Sophorolipid and Alkane Metabolism from Starmerella bombicola. *Applied and Environmental Microbiology* **80**, 766-776 (2014).

57 Cheong, S., Clomburg, J.M. & Gonzalez, R. Energy- and carbon-efficient synthesis of functionalized small molecules in bacteria using non-decarboxylative Claisen condensation reactions. *Nat Biotechnol* **34**, 556-561 (2016).

58 Durairaj, P. et al. Fungal cytochrome P450 monooxygenases of Fusarium oxysporum for the synthesis of omega-hydroxy fatty acids in engineered Saccharomyces cerevisiae. *Microb Cell Fact* **14**, 45 (2015).

59 Gatter, M. et al. A newly identified fatty alcohol oxidase gene is mainly responsible for the oxidation of long-chain omega-hydroxy fatty acids in Yarrowia lipolytica. *FEMS Yeast Res* **14**, 858-72 (2014).

60 Bowen, C.H., Bonin, J., Kogler, A., Barba-Ostria, C. & Zhang, F.Z. Engineering Escherichia coli for Conversion of Glucose to Medium-Chain omega-Hydroxy Fatty Acids and alpha,omega-Dicarboxylic Acids. *ACS Synth Biol* **5**, 200-206 (2016).

61 Seo, J.H., Lee, S.M., Lee, J. & Park, J.B. Adding value to plant oils and fatty acids: Biological transformation of fatty acids into omega-hydroxycarboxylic, alpha,omega-dicarboxylic, and omega-aminocarboxylic acids. *Journal of Biotechnology* **216**, 158-66 (2015).

62 Haushalter, R.W. et al. Production of Odd-Carbon Dicarboxylic Acids in Escherichia coli Using an Engineered Biotin-Fatty Acid Biosynthetic Pathway. *J Am Chem Soc* **139**, 4615-4618 (2017).

63 Fillet, S. & Adrio, J.L. Microbial production of fatty alcohols. *World J Microbiol Biotechnol* **32**, 152 (2016).

64 Youngquist, J.T. et al. Production of medium chain length fatty alcohols from glucose in Escherichia coli. *Metab Eng* **20**, 177-86 (2013).

65 Sheng, J., Stevens, J. & Feng, X. Pathway Compartmentalization in Peroxisome of Saccharomyces cerevisiae to Produce Versatile Medium Chain Fatty Alcohols. *Sci Rep* **6**, 26884 (2016).

66 Cao, Y.X. et al. Biosynthesis of odd-chain fatty alcohols in Escherichia coli. *Metab Eng* **29**, 113-123 (2015).

67 Lin, F.M., Marsh, E.N.G. & Lin, X.X.N. Recent progress in hydrocarbon biofuel synthesis: Pathways and enzymes. *Chinese Chemical Letters* **26**, 431-434 (2015).

68 Bertram, J.H. et al. Five Fatty Aldehyde Dehydrogenase Enzymes from Marinobacter and Acinetobacter spp. and Structural Insights into the Aldehyde Binding Pocket. *Appl Environ Microbiol* **83** (2017).

69 Liu, Y. et al. High production of fatty alcohols in Escherichia coli with fatty acid starvation. *Microb Cell Fact* **15**, 129 (2016).

70 Foo, J.L. & Leong, S.S. Directed evolution of an E. coli inner membrane transporter for improved efflux of biofuel molecules. *Biotechnol Biofuels* **6**, 81 (2013).

71 Fujita, K., Matsuyama, A., Kobayashi, Y. & Iwahashi, H. Comprehensive gene expression analysis of the response to straight-chain alcohols in Saccharomyces cerevisiae using cDNA microarray. *J Appl Microbiol* **97**, 57-67 (2004).

72 Nie, L., Ren, Y. & Schulz, H. Identification and characterization of Escherichia coli thioesterase III that functions in fatty acid beta-oxidation. *Biochemistry* **47**, 7744-7751 (2008).

73 Zhang, F.Z., Carothers, J.M. & Keasling, J.D. Design of a dynamic sensor-regulator system for production of chemicals and fuels derived from fatty acids. *Nat Biotechnol* **30**, 354-U166 (2012).

74 Mukherjee, K., Bhattacharyya, S. & Peralta-Yahya, P. GPCR-Based Chemical Biosensors for Medium-Chain Fatty Acids. *ACS Synth Biol* **4**, 1261-1269 (2015).

75 Jaspers, M.C., Meier, C., Zehnder, A.J., Harms, H. & van der Meer, J.R. Measuring mass transfer processes of octane with the help of an alkSalkB::gfp-tagged Escherichia coli. *Environ Microbiol* **3**, 512-24 (2001).

76 Gajewski, J. et al. Engineering fatty acid synthases for directed polyketide production. *Nat Chem Biol* **13**, 363-365 (2017).

77 Kovacic, F. et al. Structural and functional characterisation of TesA - a novel lysophospholipase A from Pseudomonas aeruginosa. *PLoS One* **8**, e69125 (2013).

78 Notonier, S., Gricman, L., Pleiss, J. & Hauer, B. Semirational Protein Engineering of CYP153AM.aq. -CPRBM3 for Efficient Terminal Hydroxylation of Short- to Long-Chain Fatty Acids. *Chembiochem* **17**, 1550-7 (2016).

79 Hoffmann, S.M. et al. Structure-Guided Redesign of CYP153A(M.aq) for the Improved Terminal Hydroxylation of Fatty Acids. *Chemcatchem* **8**, 3234-3239 (2016).

CHAPTER 3. MATCHING PROTEIN INTERFACES FOR IMPROVED MEDIUM-CHAIN FATTY ACID PRODUCTION

3.1 Abstract

Medium-chain fatty acids (MCFAs) are key intermediates in the synthesis of medium-chain chemicals including α-olefins and dicarboxylic acids. In bacteria, microbial production of MCFA is limited by the activity and product profile of acyl-ACP thioesterases. Here, we engineer a heterologous bacterial acyl-ACP thioesterase for improved MCFA production in *Escherichia coli*. Electrostatically matching the interface between the heterologous medium-chain *Acinetobacter baylyi* acyl-ACP thioesterase (AbTE) and the endogenous *E. coli* ACP by replacing small nonpolar amino acids on the AbTE surface for positively charged ones increased secreted MCFA titers more than three-fold. Replacing the same small nonpolar amino acids with negatively charged ones reduces MCFA titers by half. After extending the cultivation time to 72 hours, the best AbTE mutant produced 131 mg/L of MCFAs, and MCFAs are 80% of all secreted fatty acid chain lengths. Finally, a previously developed G-protein coupled receptor (GPCR)-based MCFA sensor was used to detect MCFA levels secreted by *E. coli* expressing the AbTE variants. This work demonstrates that engineering the interface between heterologous proteins and interacting host proteins is a promising approach to increase the titers of microbially-produced chemicals. Further, this work validates the GPCR-based MCFA sensor for detection of chemicals in the producer microbe's supernatant setting the stage for the sensor-guided engineering of MCFA producing microbes.

3.2 Introduction

Medium-chain fatty acids (MCFAs, C8-C12) have uses as antimicrobials and emulsifying agents and their derivatization using chain-tailoring enzymes results in an array of chemicals including alkenes, α-olefins, esters, ω-hydroxy-carboxylic acids, α,ω-dicarboxylic acids, alcohols, and ketones. Microbially, MCFAs are generated via hydrolysis of medium-chain fatty acyl-ACPs obtained from Type II fatty acid synthase (FAS)[1], or medium-chain fatty acyl CoAs produced by Type I FAS[2] or the reverse β-oxidation pathway[3]. Hydrolysis of fatty acyl intermediates is performed by medium-chain acyl-ACP thioesterases (TEs) from bacteria, higher plants (FatBs), and mammals, or medium-chain acyl-CoA TEs from bacteria. While bacterial acyl-ACP TEs tend to hydrolyze C8-C16 acyl-ACPs[4], such as *Acinetobacter baylyi* TE[5], FatBs have a narrower substrate profile, such as *Cuphea palustris* TE (C8:0 acyl-ACPs)[6], and *Umbellularia californica* TE (C12:0 acyl-ACPs)[7]. The rat acyl-ACP TE has been shown to produce primarily C6-C8 fatty acids[8]. On the acyl-CoA front, *E. coli* ydiI hydrolyzes C6-C10 acyl-CoAs[4].

Microbes do not naturally synthesize MCFAs extensively, and microbial MCFA production is hindered by the inefficient expression of heterologous plant thioesterases and the broad substrate profile of bacterial thioesterases. Long chain acyl-ACP TEs have been engineered for improved MCFA production via active site mutagenesis and computationally-guided approaches. For example, *E. coli* 'TesA:L109P (signal peptide removed) preferentially hydrolyzes 12:0 and C14:0 acyl-ACPs, unlike the wild type's preference for C16:0 acyl-ACPs[9]. Similarly, *Pseudomonas aeruginosa* TesA:D17S/L462R has improved activity on C12:0 acyl-CoAs[10]. More recently, using an

iterative protein redesign and optimization algorithm 'TesA sequences that preferentially bind C8-C12 substrates were selected leading to 'TesA:S122K/Y145K/L146K and 'TesA:M141L/Y145K/L146K, which had a 1.8-fold improvement in C12 mole fraction and a 10-fold improvement in C8 mole fraction over wild type, respectively[11].

Using a medium-chain acyl-ACP TE as the engineering starting point, however, has the potential to result in a TE variant with an almost exclusive MCFA product profile, potentially leading to higher MCFA yields. Engineering medium-chain acyl-ACP TEs has a unique set of challenges. First, medium-chain acyl-ACP-TEs are heterologous to *E. coli* and may have problems interfacing with the host machinery. Second, active site engineering of medium-chain acyl-ACP TEs may not prove to be as fruitful, as mutations may need to be much more subtle, potentially in the second shell, and overall more difficult to identify. We hypothesized that engineering the interface of a heterologous medium chain TE to better complement the surface of *E. coli* ACP may improve MCFA titers (Figure 3.1). In Type II FAS, ACP is bound to the fatty acyl-chain and interacts with all the proteins in fatty acid biosynthesis. The fatty acyl chain is buried in the ACP hydrophobic core and protein-protein interactions between the ACP and partner enzymes release the acyl chain from the ACP core and into the partner enzyme's active site[12]. It is the tight binding of ACP to its protein partners, such as FAS subunits, that enables efficient fatty acid biosynthesis[13]. Indeed, crosslinking studies show that *E. coli* Type II FAS ACP binds more tightly to its cognate *E. coli* Type II FAS ketoacid synthase (KS) than *Streptomyces maritimus* Type II polyketide synthase KS[13].

Here, we engineer the medium-chain acyl-ACP TE from *Acinetobacter baylyi* (AbTE) to better interface with *E. coli* ACP to improve MCFA production. First, we

docked *E. coli* ACP with the endogenous *E. coli* TE ‘TesA and identify potential contact residues involved in stabilizing the ACP-‘TesA interaction. Next, we mutated the equivalent positions in AbTE to the amino acids found in *E. coli* ‘TesA and measured its fatty acid profile. We find that mutation of just two residues on the AbTE surface, G17 and A165 to arginines, improves MCFA titers more than 3-fold when compared to expression of AbTE wild type in *E. coli*. This work demonstrates that engineering the interface of heterologous enzymes to better couple with endogenous host enzymes may be a useful strategy to improve the microbial production of chemicals that require the expression of heterologous enzymes. Improving the microbial production of MCFAs is significant because MCFAs are key intermediates in the biosynthesis of medium-chain chemicals, including α-olefins, dicarboxylic acids and hydroxyacids, which are important targets in the polymer industry. Finally, using a previously developed G-protein coupled receptor (GPCR)-based MCFA sensor[14], we detect different MCFA levels of produced by *E. coli* expressing the AbTE mutants. Therefore, this work also sets the stage for future sensor-guided engineering of MCFA producing microbes.

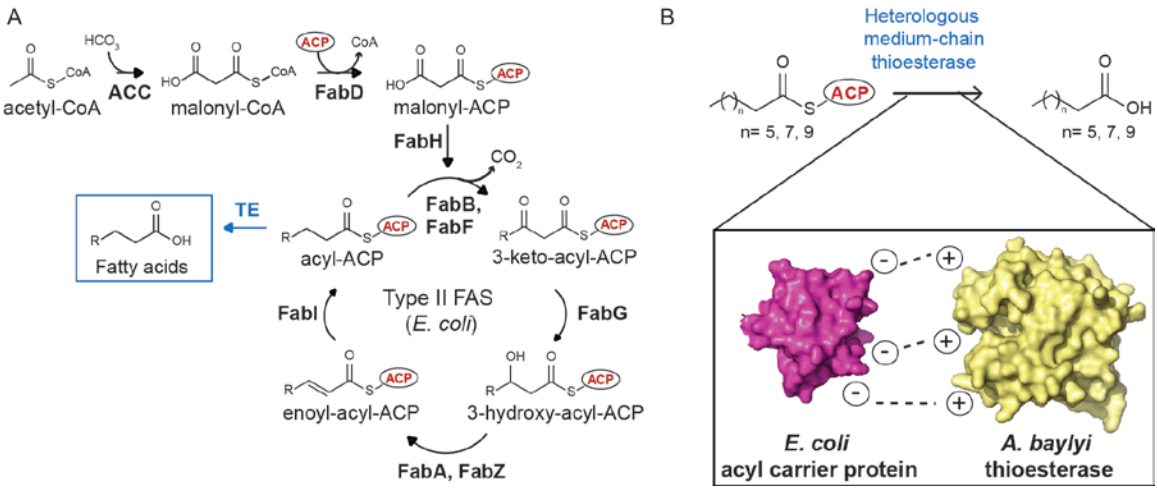

Figure 3.1: Medium-chain fatty acid biosynthesis in *Escherichia coli*. A. *E. coli* Type II fatty acid synthase (FAS) extends and reduces an acyl chain bound to acyl-carrier protein (ACP). All enzymes in FAS interact with ACP. Thioesterases (TEs) hydrolyze acyl-ACPs to free fatty acids of different chain lengths according to their substrate specificity. B. Matching the surface interface between the native *E. coli* ACP and medium-chain heterologous thioesterases, such as *Acinetobacter baylyi* thioesterase, improves medium-chain fatty acid production in *E. coli*.

3.3 Results

3.3.1 Screening medium-chain acyl-ACP TEs and E. coli hosts for MCFA production

To identify the acyl-ACP TE that results in the highest secreted MCFA titers, we expressed the bacterial thioesterase from *Acinetobacter baylyi*[5], and the plant thioesterases from *Cocos nucifera*[6], *Cuphea palustris*[6], and *Umbellularia californica*[15] in *E. coli* MG1655 (Figure 3.2). The percent sequence identity of these TEs to one another ranges from 15-17%; *A. baylyi* TE has the highest percent identity with *E. coli* 'TesA at 38% (Figure 3.3). We measured secreted fatty acids as they could be continuously extracted from the culture broth, overcoming issues with cell lysis, and potentially reducing the overall cost for MCFA production. AbTE results in the highest secreted MCFA titers at 29 mg/L, consisting of octanoic, decanoic and dodecanoic acid. *C*

nucifera TE produced only dodecanoic acid at 5 mg/L, while *C. palustris* TE produced mostly octanoic acid at 8 mg/L. *U. californica* TE produced a mixture of decanoic and dodecanoic acid at 3 mg/L and 12 mg/L respectively. Given that AbTE resulted in the highest MCFA titers, we decided to engineer this enzyme.

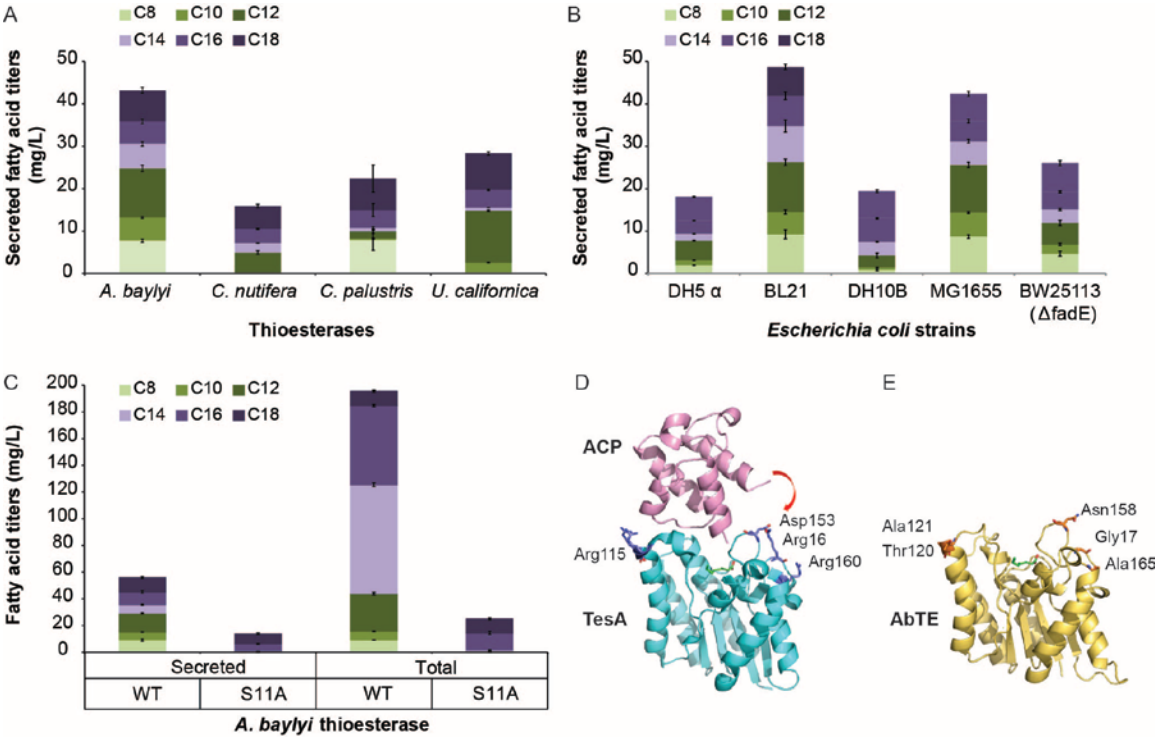

Figure 3.2: Medium-chain fatty acid production and calculated *E. coli* ACP interphase with thioesterases. A. Secreted fatty acid titers of *E. coli* (MG1655) expressing four heterologous thioesterases. B. Secreted fatty acid titers by different *E. coli* strains expressing *Acinetobacter baylyi* thioesterase (AbTE wt). C. Secreted and total (secreted plus intracellular and membrane bound) fatty acid titers produced by *E. coli* expressing AbTE:WT and inactive AbTE (AbTE:S11A). The experiments were done in triplicate and the error bars represent the standard deviation from the mean. D. Docking of *E.coli* ACP (magenta, PDB ID: 1FAE) and *E. coli* 'TesA (cyan, PDB ID 1U8U) identifies potential residues on the 'TesA surface that are important for interactions between 'TesA and ACP. E. Homology model of AbTE with surface residues equivalent to 'TesA labelled.

The *E. coli* genomic background has been shown to affect chemical production[16]. We expressed AbTE in five different *E. coli* hosts: DH5α, BL21, DH10B, MG1655, and

BW25113 Δ*fadE*, and measured secreted MCFA titers. We included the *fadE* deletion as it has been shown to improve fatty acid production in *E. coli*[17]. Surprisingly, *E. coli* hosts BL21 and MG1655 resulted in the highest MCFA productions at 26 mg/L and BW25113 Δ*fadE* produced only 12 mg/L (Figure 3.2). Based on these results, we moved forward with AbTE expressed in *E. coli* MG1655.

	E. coli tesA	*A. baylyi* TE
E. coli tesA		38.3%
A. baylyi TE	38.3%	
C. nutifera TE	19.0%	16.9%
U. californica TE	21.1%	14.7%
C. palustris TE	16.7	16.9%

Figure 3.3: Percent sequence identity between thioesterases used in this work. Bacterial thioesterases (blue). Plant thioesterases (green).

3.3.2 Engineering AbTE for improved MCFA titers

Expression of the non-functional AbTE:S11A in *E. coli* produces only saturated long-chain (C14-C18) fatty acids (LCFA) due to the presence of endogenous long chain thioesterases in *E. coli*. Expression of AbTE wild type (AbTE:WT) in *E. coli* produced ~29 mg/L of secreted MCFAs or ~52% of all secreted fatty acid chain lengths. When total fatty acids were measured, i.e. secreted fatty acids plus intracellular and membrane bound fatty acids, AbTE:WT expressed in *E. coli* produced ~48 mg/L of MCFAs, which is ~22% of total fatty acid chain lengths. While MCFA levels increased by 65% when taking into account intracellular and membrane bound fatty acids, LCFA levels increase more than 6-fold. Specifically, AbTE:WT produced octanoic, decanoic, and dodecanoic

acid at 9 mg/L, 6, mg/L and 14 mg/L, respectively (Figure 3.2). In addition to saturated fatty acids, AbTE:WT also produced small levels of unsaturated C12-C16 fatty acids (Figure 3.4).

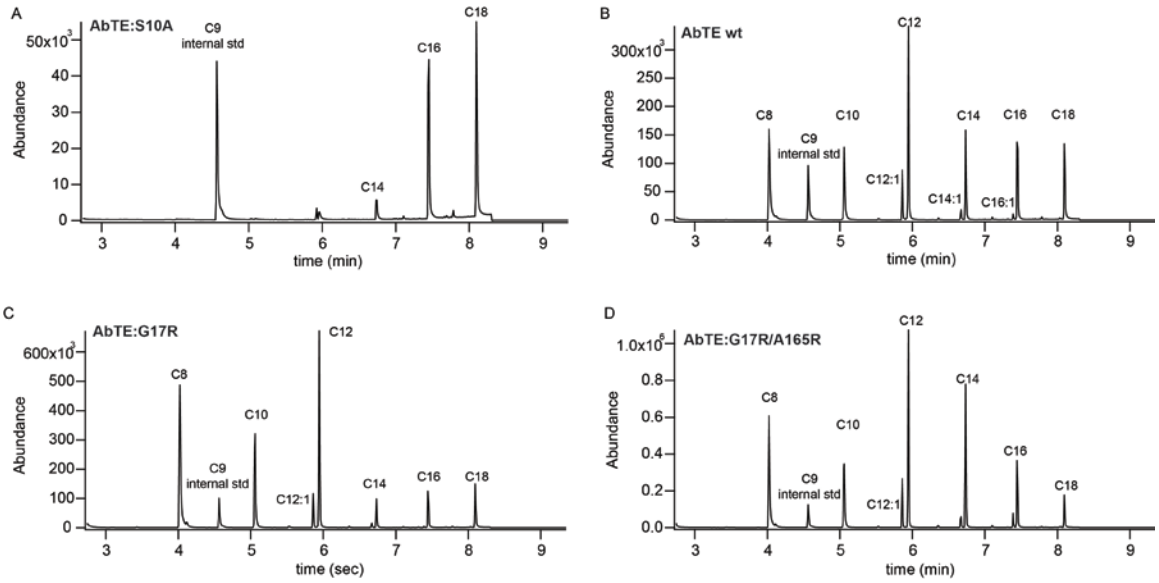

Figure 3.4: Gas chromatograms of AbTE and variants. A. AbTE:S11A (inactive enzyme), B. Wild-type AbTE) C. AbTE:G17R D. AbTE:G17R/A165R. Single Ion Monitoring: 74 and 87.

To identify the interface between *E. coli* 'TesA and *E. coli* ACP we used ClusPro[18], which takes into account only the 'TesA-ACP protein interactions to dock *E. coli* ACP (PDB ID: 2FAE) and *E. coli* 'TesA bound to octanoic acid (PDB ID: 1U8U) (Figure 3.2). Using the model, we identified eight positions on 'TesA that are potentially part of the ACP-'TesA interface: Y15, R16, R77, N112, R115, R116, D153, and R160. Structural alignment of 'TesA with a AbTE homology[19] model revealed that all positions except for R16 (AbTE: G17), R115 (AbTE: T120), R116 (AbTE: A121), D153 (AbTE: N158) and R160 (AbTE: A165) had the same amino acids in these two proteins (Figure 3.2, Figure 3.5). Interestingly, four of the five amino acids that are different between 'TesA and

AbTE are positively charged arginines, which could help stabilize the 'TesA-ACP interaction as ACP has a highly negative surface. We hypothesized we could replicate these interactions between AbTE and ACP to improve MCFA production.

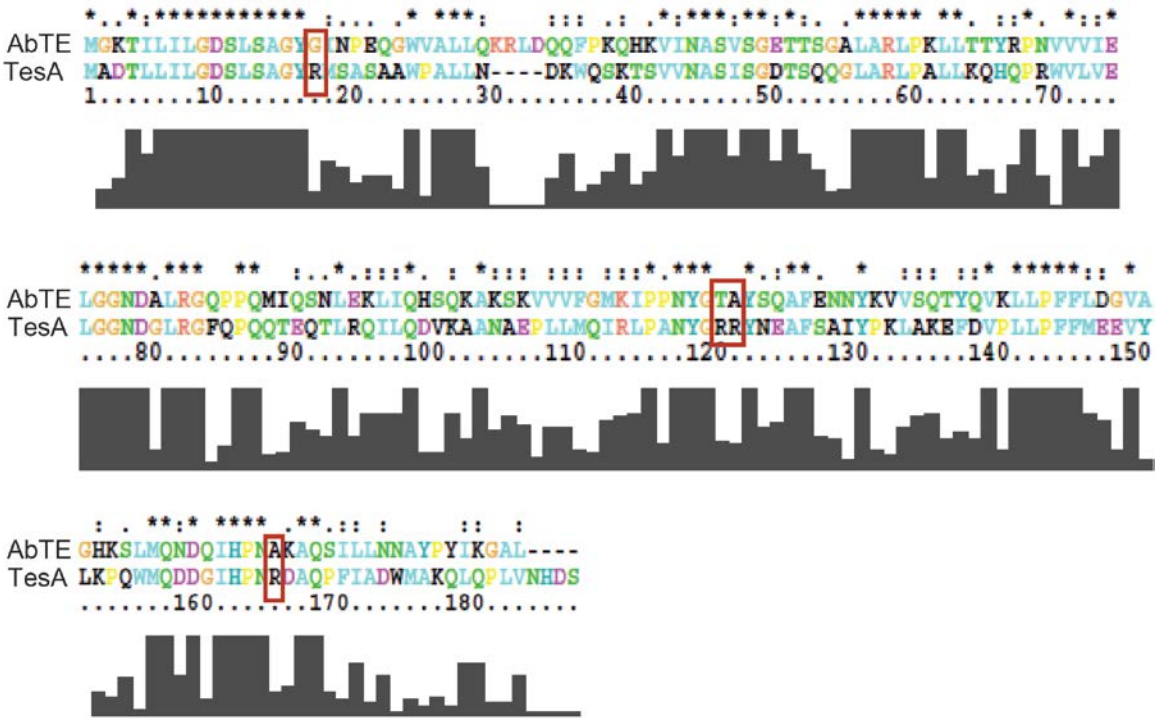

Figure 3.5: Amino acid sequence alignment of AbTE and *E. coli* 'tesA. Red boxes indicate amino acids targeted for mutagenesis.

We mutated positions 17, 120, 121, and 165 on AbTE to arginines to generate AbTE:G17R, AbTE:T120R, AbTE: A121R and AbTEA165R, and measured their secreted fatty acid titers (Figure 3.6, Figure 3.4). Expression of AbTE:G17R in *E. coli* resulted in ~76mg/L of secreted MCFAs, more than double the secreted MCFA titers from AbTE:WT. The MCFAs produced by AbTE:G17R accounted for ~74% of secreted fatty acids of all chain lengths. In particular, octanoic, decanoic and dodecanoic acid were produced at 30 mg/L, 18 mg/L and 28 mg/L, respectively. Expression of AbTE:A165R in *E. coli* resulted in slightly lower secreted MCFA titers than AbTE:WT. To determine if

the effects of these mutations on MCFA titers were additive, we constructed all double mutants using AbTE:G17R as the scaffold. Expression of AbTE:G17R/A165R in *E. coli* resulted in ~98 mg/L of secreted MCFAs. This is a 29% increase in secreted MCFA titers when compared to expression of AbTE:G17R and a more than 3-fold in secreted MCFA titers when compared to AbTE:WT. Although AbTE:G17R/A165R secreted MCFA titers improved, the MCFA percentage was only 55% of secreted fatty acids of all chain lengths. The major constituent in the AbTE:G17R/A165R secreted MCFA profile was dodecanoic acid at 45 mg/L. The increase in secreted MCFA titer achieved by AbTE:G17R/A165R was unexpected as AbTE:A165R resulted in the lowest MCFA titers from all single mutants. Finally, we generated the triple mutants using AbTE:G17R/A165R as the scaffold. At this stage, we also mutated the fifth position on AbTE that varies from ‘TesA, AbTE:N158D. This fifth position does not change the amino acid to a positively charged arginine, but to a negatively charged aspartate. Nevertheless, the aspartate could still form part of a stabilizing interaction. As Figure 3C shows, none of the triple mutants resulted in improved secreted MCFA titers. Of note, positions A121 and T120 are located on the other side of the AbTE binding pocket than G17 and A165. For completion, we also generated the remaining double mutants AbTE:T120R/A121R, AbTE:T120R/A121R, and AbTE:A121R/A165R and measured their secreted fatty acid titers (Figure 3.7). While two of these double mutants produced comparable secreted MCFA titers to AbTE:WT, AbTE:T120R/A121R produced ~45 mg/L of secreted MCFAs. Taken together, AbTE:G17R/A165R results in the highest secreted MCFA titers (98 mg/L), yet AbTE:G17R has the highest percentage of secreted

MCFAs (74%), a trend that holds true whether analyzing secreted or total fatty acids (Table 3-1).

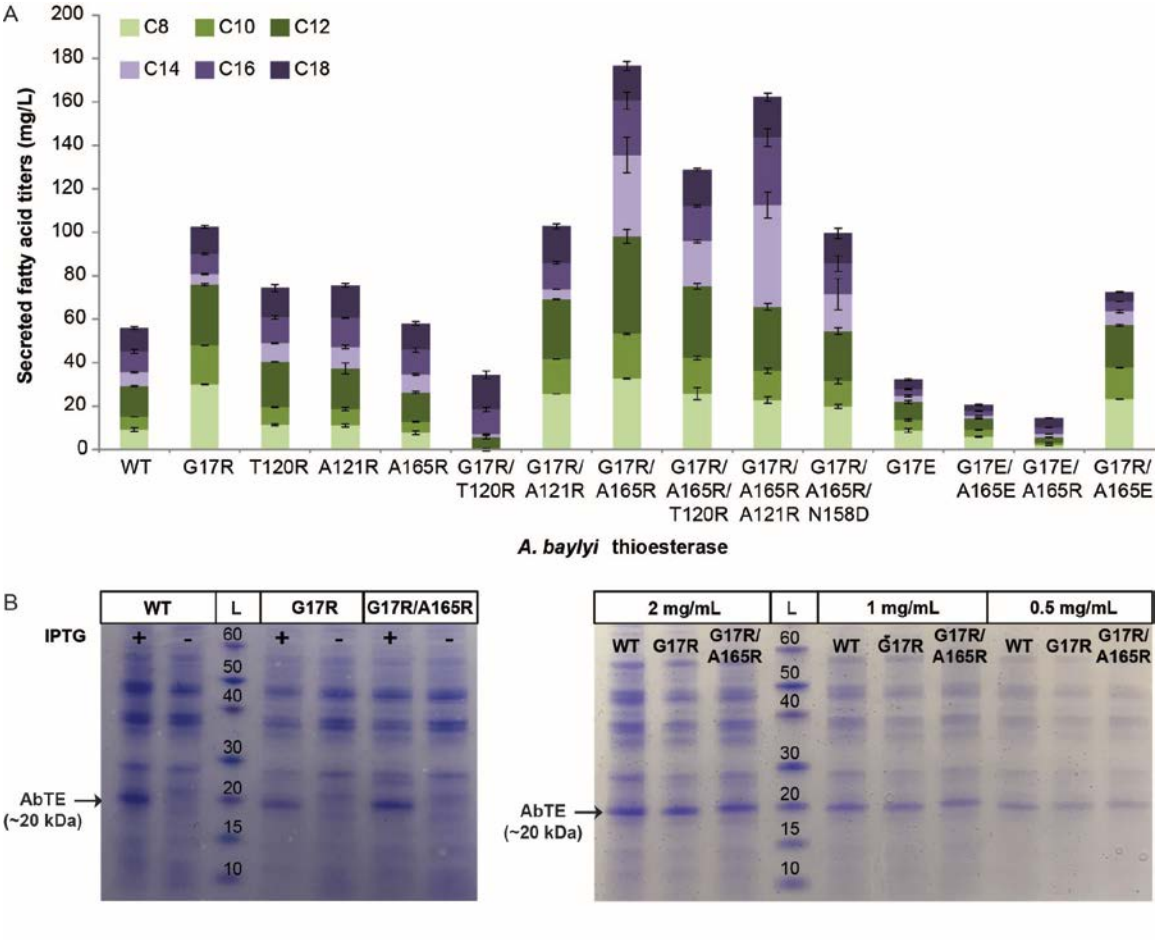

Figure 3.6: Secreted fatty acid and protein levels of *E. coli* expressing *Acinetobacter baylyi* thioesterase (AbTE) and AbTE mutants. A. Secreted fatty acid levels of *E .coli* expressing AbTE single, double, and triple arginine mutants, as well as AbTE single and double glutamate mutants. All experiments were done in triplicate and the error bars represent the standard deviation from the mean. B. Left: Coomassie stained SDS-PAGE gel of induced (+IPTG) and uninduced (-IPTG) *E. coli* cultures expressing AbTE:WT, AbTE:G17R and AbTE:G17R/A165R. Right: Serial dilution of induced *E. coli* cultures expressing AbTE:WT, AbTE:G17R and AbTE:G17R/A165R.

It is possible that the improved MCFA profile of AbTE:G17R and AbTE:G17R/A165R when compared to AbTE:WT is the result of the mutants' superior

expression in *E. coli*. Positions 17 and 165 are located on the AbTE surface and mutating small hydrophobic amino acids to positively charged arginines may improve solubility. A SDS-PAGE gel of *E. coli* expressing AbTE:WT, AbTE:G17R or AbTE:G17R/A165R showed comparable soluble expression of the three enzymes over 3 dilutions of cell lysate (Figure 3.6).

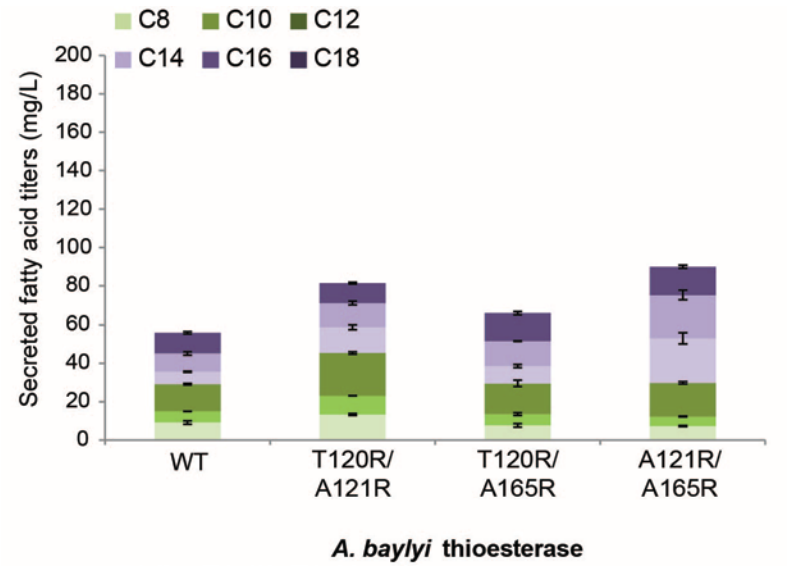

Figure 3.7: Remaining AbTE double mutant combinations. Secreted fatty acid titers produced by *E. coli* expressing AbTE, AbTE:T120R/A121R, AbTE:T120R/A165R, and AbTE:A121R/A165R. All experiments were done in triplicate and the error bars represent the standard deviation from the mean.

In *E. coli*, arginines on the surface of FAS enzymes are essential for interaction with ACP, which has a highly negatively charged surface[20]. Mutation of these arginines in FAS enzymes to negatively charged amino acids results in decreased interactions with ACP (FabA), decreased specific activity with ACP substrates (FabH, FabG), and decrease K_{cat} and increased K_M with ACP substrates (FabI) [21-24]. If the newly introduced arginines on AbTE help stabilize AbTE-ACP interactions, mutating AbTE positions 17 and 165 to a negatively charged amino acids, such as glutamate, should disrupt ACP

interactions resulting in lower MCFA titers. *E. coli* expression of both the single AbTE:G17E and AbTE:G17E/A165E result in overall lower secreted fatty acid titers than AbTE:WT (Figure 3.6). AbTE:G17E and AbTE:G17E/A165E results in ~21mg/L and 14 mg/L of secreted MCFAs, respectively, which is ~69% of secreted fatty acids of all chain lengths. Interestingly, AbTE:G17R/A165E results in ~57 mg/L of secreted MCFAs, which is ~79% of secreted fatty acids of all chain lengths. This favorable effect in secreted MCFAs or secreted fatty acids of all chain lengths is not seen in AbTE:G17E/A165R, making position 17 key in stabilizing AbTE-ACP interactions. Taken together, the arginines on the AbTE surface do not affect the protein expression in *E. coli*, and enhance AbTE's interaction with *E. coli* ACP.

Table 3-1: Saturated fatty acid percent composition produced by AbTE:WT and variants.

	Thioesterase	***E. coli* strain**	**Media**	**C8:0 (%)**	**C10:0 (%)**	**C12:0 (%)**	**C14:0 (%)**	**C16:0 (%)**	**C18:0 (%)**
Supernatant	AbTE:WT	MG1655	M9	16.5	10.5	25.2	11.5	17.1	19.2
	AbTE:G17R	MG1655	M9	29.3	17.5	27.3	4.7	9.1	12.1
	AbTE:G17R/A165R	MG1655	M9	18.5	11.7	25.4	21.2	14.2	9
Total	AbTE:WT	MG1655	M9	4.7	3.3	14.4	41.7	30.4	5.5
	AbTE:G17R	MG1655	M9	16.1	9.9	25.0	29.4	13.9	5.7
	AbTE:G17R/A165R	MG1655	M9	9.4	6.0	18.8	38.7	18.6	8.5

3.3.3 Extending cultivation time to increase MCFA titers

Expecting to accumulate higher MCFA titers during a longer cultivation period, we increased the cultivation time from 24hrs to 72hrs (Figure 3.8). When expressing AbTE:WT in *E. coli*, secreted MCFA titers almost tripled between 24hrs and 72hrs from ~29 mg/L to ~79 mg/L, jumping from being 52% to 69% of all secreted fatty acid chain lengths. We speculate that LCFAs are more likely to be ligate to CoA by endogenous FadD and enter the β-oxidation pathway[25], thus leading to a reduction of LCMA levels over time. Interestingly, cultivating *E. coli* expressing AbTE:G17R for 72hrs did not result in any changes in MCFA titers. When expressing AbTE:G17R/A165R in *E. coli*, secreted MCFAs increased from 98 mg/L to 131 mg/L when extending the cultivation form 24hrs to 72 hrs. This increasing in MCFA titers increased the percentage of secreted MCFAs from 55% to 80%. It is possible that the low increase in MCFA titers observed with the AbTE mutants between 24hrs and 72hrs is due to lost in activity after 24hrs. It has been shown that adding solubility tags to heterologous proteins increases their viability inside the cell[26]. Attaching maltose binding protein (MBP) to the N-terminus of the AbTE mutants, the terminus at the opposite end of the ACP-AbTE interface, results only in LCFAs and no MCFAs were detected (Figure 3.10).

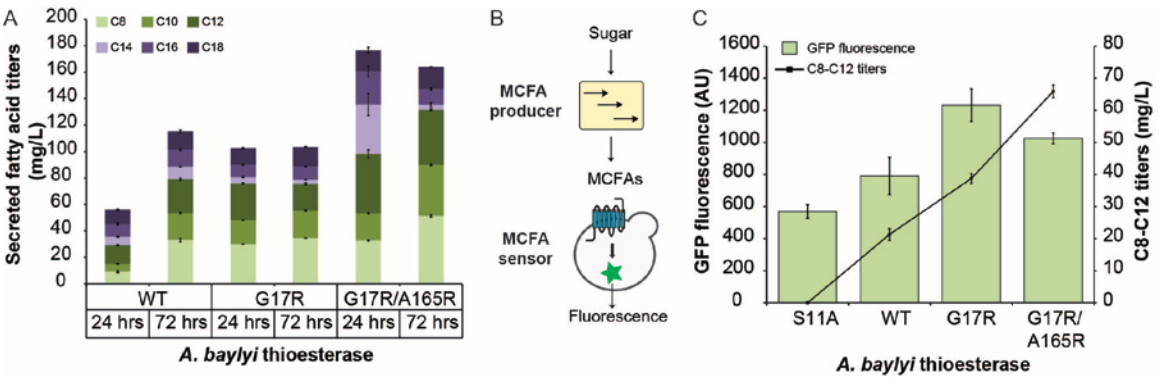

Figure 3.8: Cultivation time effect on medium-chain fatty (MCFA) acid titers and MCFA detection in the producer supernatant via a GPCR-based MCFA sensor. A. Time course of saturated fatty acid titers produced by *E. coli* expressing AbTE, AbTE:G17R, and AbTE:G17R/A165R. B. Schematic of E. coli produced MCFA by a GPCR-based MCFA sensor. MCFA producer: glucose is fed to *E. coli* expressing a AbTE to secrete MCFAs to the supernatant. MCFA sensor: *S. cerevisiae* expressing the GPCR OR1G1 (blue) detects MCFAs in the producer supernatant resulting in expression of green fluorescent protein. D. Detection of MCFAs by the MCFA sensor in the E. coli supernatant. Sensor GFP fluorescence and MCFA (C8-C12) titers as a function of AbTE variant expressed in *E. coli*.

3.3.4 Detecting MCFAs secreted by engineered E. coli using a GPCR-based MCFA sensor

Recently, we developed a GPCR-based MCFA sensor in yeast by expressing a mammalian olfactory receptor known to bind MCFAs, OR1G1, and coupling it to the yeast mating pathway, resulting in green fluorescent protein expression upon MCFA detection on the cell surface[14] (Figure 3.9).

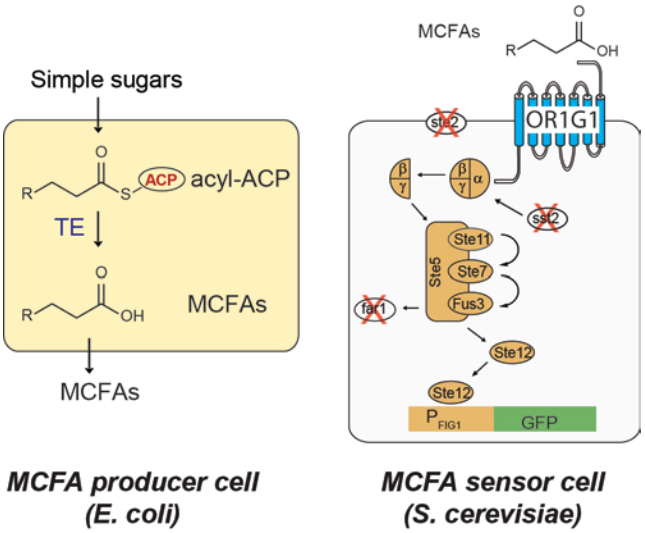

Figure 3.9: MCFA *E. coli* producer cell and *S. cerevisiae* sensor pair. MCFAs produced by the producer are detected by the sensor via a fluorescent readout.

Here, we set out to determine the extent to which the MCFA sensor can detect different levels of microbially produced MCFAs secreted by a chemical producer microbe (Figure 3.8). The proposed MCFA detection has the added difficulty that the sensor cell (*Saccharomyces cerevisiae)* is a different species than the chemical producer microbe (*E. coli*). Further, since we envision using the MCFA sensor to screen MCFA secreted by producer microbes in a medium-throughput fashion, the MCFA production was carried out in 96-deep well plates rather than 5mL test tubes, resulting in slightly lower secreted MCFA titers, likely due to poor aeration. To detect MCFA secreted by *E. coli* with the *S. cerevisiae* MCFA sensor we 1) lowered the *E. coli* supernatant pH from 7 to 4 as the MCFA sensor detects the protonated form of the MCFAs (pKa =4.5), and 2) diluted the *E. coli* supernatant 2-fold in fresh yeast media so that the MCFA titers fell within the linear range of the sensor (C8: 2.7-36 mg/L; C10: 6-43 mg/L; C12: 0.2-50 mg/L). As Figure 3.8 shows, we see a linear correlation between MCFA sensor fluorescence and the MCFA titers secreted by *E. coli* expressing the different AbTE

variants. The sensor fluorescent signal saturates between AbTE:G17R (C8: 24 mg/L; C10:11 mg/L; C12: 4 mg/L) and AbTE:G17R/A165R (C8: 34 mg/L; C10:19 mg/L; C12: 13 mg/L) even though concentrations of each, C8, C10 and C12 produced by AbTE:G17R and AbTE:G17R/A165R fall within the MCFA sensor range. The sensor is likely detecting the contribution of all three MCFAs together leading to early signal saturation.

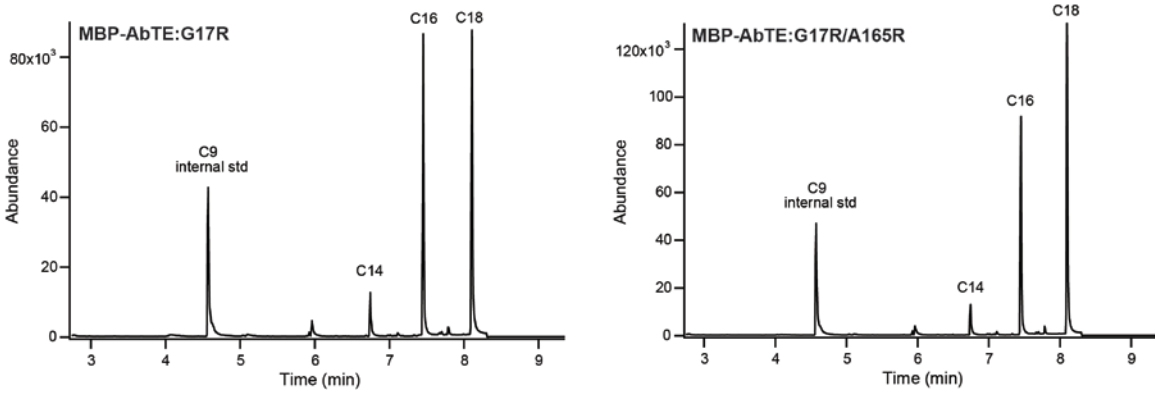

Figure 3.10: Effect of fusing maltose binding protein (MBP) to AbTE mutants. Gas chromatograms of MBP-AbTE:G17R (left) and MBP-AbTE:G17R/A165R (right). Single Ion Monitoring: 74 and 87.

3.4 Discussion

We engineered the surface of a heterologous enzyme (thioesterase) to better couple to the endogenous *E. coli* enzyme (ACP) to increase chemical titers (MCFAs). Replacement of two small nonpolar residues on the AbTE surface predicted to contact *E. coli* ACP, which has a highly negatively charged surface, with positively charged arginines, the amino acid found at the equivalent positions in *E. coli* 'TesA, resulted in more than 3-fold improvement in secreted MCFA titers. Replacing the small nonpolar residues of the AbTE surface with negatively charged glutamate resulted in lower overall secreted fatty

acid titers. We speculate that improving the interface of AbTE and *E. coli* ACP enables AbTE to more efficiently accept medium chain fatty acyl-ACPs, thus improving MCFA titers. As shown, the improvement in MCFA titers is not grounded on improved expression of the AbTE mutants when compared to AbTE:WT but on the favorable electrostatics between the AbTE mutants and *E. coli* ACP.

In the future, engineering the interface of heterologous proteins to better match the interface of production host enzymes could be applied to more distantly related proteins, such as plant thioesterases. Such an approach may prove even more beneficial for more distantly related enzymes that may only marginally interact with endogenous host proteins. A key limitation will be the low sequence homology between heterologous and endogenous proteins, making it difficult to identify key residues for mutagenesis (Figure 3.3). Crystal structures of such enzymes (e.g. plant thioesterase) would enable structural alignments and identification of such residues. In addition to MCFA production, the matching interface strategy could also be applied to medium-chain methyl ketone production by better coupling *E. coli* ACPs to heterologous β-ketoacyl-ACP thioesterases, such one from *Solanum habrochaites* (ShMKS2)[27]. .

The GPCR-based MCFA sensor (*S. cerevisiae*) detected MCFAs secreted by a producer microbe from a different species (*E. coli*) setting the stage for the sensor-guided engineering of MCFA producing *E. coli*. Moreover, as the GPCR-based MCFA sensor appears to be producer microbe agnostic, secreted MCFA levels from other industrially relevant hosts, such as cyanobacteria and algae, could also be targeted for sensor-guided engineering. Here, we show that MCFA detection can occur in 96-well plates enabling future medium-throughput (10^3 samples per day) screening of MCFA-producing

microbes. This screening throughput is one order of magnitude faster than current state of the art for the detection of microbially produced MCFA, i.e. gas chromatography/mass spectrometry (GC-MS). The MCFA sensor could be used to screen large libraries of MCFA producing microbes, to down select key mutants for testing using GC-MS. The medium-throughput screening of microbial MCFAs would be especially valuable for engineering of plant thioesterases that lack crystal structures and have very low sequence identity to bacterial thioesterases.

3.5 Materials and Methods

3.5.1 Reagents

Chemical standards were purchased from: C8:0 (A10991, Alfa Aesar); C10:0 (A15658, Alfa Aesar); C12:0 (A12492, Alfa Aesar); C14:0 (A10257, Alfa Aesar); C16:0 (P0006, TCI); C18:0 (S0080, TCI).

3.5.2 Plasmid construction

Non-codon optimized *Acinetobacter baylyi* TE (AbTE), *S. cerevisiae* codon-optimized *Cuphea palustris* TE (CpTE), *E. coli* codon-optimized *Umbellularia californica* TE (UcTE) were commercially synthesized and cloned under P_{TRC} in pMB1-P_{TRC}-AgGPPS-$(GSG)_2$-AgPS (pSS185) between *NcoI/XmaI* to generate pMB1-P_{TRC}-AbTE (pSS192), pMB1-P_{TRC}-CpTE (pSS183), and pMB1-P_{TRC}-UcTE (pSS193). *S. cerevisiae* codon-optimized CnTE was amplified from pESC-LEU2-P_{TEF1}-P_{HXT7}-CnTE (pSS81) with primers SS455/SS456 and cloned under P_{TRC} in pSS185 between *NcoI/XmaI* to generate pMB1-P_{TRC}-CnTE (pSS174).

3.5.3 AbTE mutant generation

Site-directed mutagenesis was performed using the QuikChange protocol with some modifications. PCR reaction: 0.8 ng/μL of template, 2.5 ng/μL of each primer, 1X iProof HF polymerase buffer, 0.02 U/μL iProof polymerase (BioRad), 0.5 mM dNTPs to 50 μL final volume. Details on templates and primers used are in **Table SI4.** Thermocycler protocol: 95°C 1min, 17 cycles: [95°C 50sec, 60°C 50sec, 72°C 2min 30sec], 72°C 7 min. DpnI (0.4 U/μL) was added to PCR reaction and incubated at 37°C for 1 hour and heat inactivated at 65°C for 20 min. 10 μL of reaction was transformed into competent DH10B *E. coli* cells.

3.5.4 Fatty acid production and quantification

MCFA Production: Overnight cultures of *E.coli* MG1655 expressing AbTE:WT or AbTE variants were diluted 1:50 in 5 mL of M9 media (0.5% glucose, amp^{100}) and grown at 37°C, 250 r.p.m. until reaching an OD_{600} = 0.3-0.4. The cells were then induced with 500 μM of IPTG (500 mM stock) and grown at 30°C, 250 r.p.m. for 24 or 72 hrs. *Fatty acid analysis:* For secreted fatty acids, *E. coli* cultures were vortexed for 3 sec, 600 μL of culture removed and centrifuged for 10 min at 7354*g*. Next, 400 μL of the supernatant was removed for derivatization. For total fatty acids, 400 μL of culture was used for derivatization. *Fatty acid* d*erivatization:* Fatty acids were derivatized to fatty acid methyl esters and analyzed via GC-MS as described in Torella et. al., 2013[28] with some modifications. To the 400 μL of sample, 50 μL of 10% (wt/vol) NaCl, 50 μL of glacial acetic acid, 20 μL of 90.5 mg/L nonanoic acid (internal standard), and 200 μL of ethyl acetate were added and the mixture was vortexed for 5 sec. The mixture was then

centrifuged at 12,098*g* for 10 min. Methyl esters were generated by mixing 100 μL of the ethyl acetate layer with 900 μL of a 30:1 mixture of methanol and 37% (vol/vol) HCl in a 2 mL microcentrifuge tube, vortexed for 5 sec, and incubated at 50°C for 1 hr. After cooling to room temperature, 500 μL of water and 500 μL of hexanes were added. The mixture was vortexed for 5 sec, 100μL of the hexane layer was taken and mixed with 400 μL of ethyl acetate for analysis via GC-MS. *FAME quantification*: The samples were analyzed using Agilent 7890A/Agilent 5975 MS detector using a DB-5MS column. The inlet temperature was set to 300°C, flow at 1 mL/min, the oven at 70°C for 1 min, ramp at 30°C/min to 290°C, and held for 1 min at 290°C. Standard curves of C8-C18 fully saturated FAMEs (Alfa Aesar/TCI) were used for sample quantification.

3.5.5 E. coli secreted MCFA detection by S. cerevisiae MCFA sensor

E. coli MCFA production: *E. coli* MG1655 expressing AbTE:WT (PPY1331), AbTE:S11A (PPY1332), AbTE:G17R (PPY1333) or AbTE:G17R/A165R (PPY1340) were grown overnight in 200 μL of LB amp^{100} in a 96-well plate. The overnight cultures were diluted 1:50 in 1 mL of M9 media (0.5% glucose, amp^{100}) in a 96-deep well plate. The plate was grown at 37°C, 250 r.p.m. for 3 hrs and 45 min. The cells were then induced with 500 μM of IPTG and grown at 30°C, 250 r.p.m. for 24 hrs. Next, the plate was centrifuged for 10 min at 3005*g* and 400 μL of the supernatant of each well was used for fatty acid derivatization and quantification, while 50μL was placed into a fresh 96-well plate for sensing using the MCFA sensor. *Sensor preparation*: The yeast sensor strain (PPY653) was grown overnight in synthetic complete media with 2% glucose lacking histidine and leucine (SD glu (HL^-)). The overnight culture was used to inoculate 40 mL of SD glu (HL^-) to an OD_{600} = 0.06 in a 250 mL Erlenmeyer flask and incubated

for 18 hours at 15°C (150 rpm). The cells were centrifuged for 10 min at 2301*g* and resuspended in fresh 2 mL SD glu (HL^-). *MCFA Sensing*: To the 50 µL of supernatant, 50 µL of SD glu (HL^-) and 0.5 µL of 2.6M HCl was added. Next, the sensor cells were added to the supernatant to reach an OD_{600} = 0.02. The supernatant/sensor mixture was incubated for 4 hours at 30°C (no shaking) before reading for cell fluorescence using a flow cytometer. GFP fluorescence was measured using a BD Fortessa flow cytometer with the following settings: 488nM laser line, FSC: 80 V, SSC: 135 V, FITC: 400 V. Fluorescence data was collected from 60µL of sample volume from each well with a flow rate of 1µL/sec. Flow cytometry histogram analysis was done using FlowJo software**.**

3.5.6 Homology model and docking

Phyre2 software was used to generate the AbTE homology model. The model was generated in intensive mode and the confidence in the model for 181 residues (99%) was modelled at >90% accuracy. ClusPro software was used for docking of *E. coli* acyl-carrier protein (PDB ID: 2FAE, chain A) onto *E. coli* tesA (PDB ID: 1U8U). The 9 balanced models were analyzed via PyMol and were used to deduce the hypothetical protein-protein interactions for tesA and AbTE.

3.5.7 SDS-PAGE gel

PPY1331, PPY1333, PPY1340 were grown overnight in LB media amp^{100}. The overnight cultures were diluted 1:50 in 1 mL of M9 media (0.5% glucose, amp^{100}). The cultures were grown at 37°C, 250 rpm. until they reached an OD_{600} = 0.30-0.40. The cells were then induced with 500 µM of IPTG (500 mM stock conc.) and grown at 30°C, 250 rpm for 24 hours. 1 mL of sample was removed from the culture medium and centrifuged

5 min at 7354*g*. The supernatants were discarded, and the pellet resuspended in 200 μL of 1X PBS. The cell suspension was sonicated 2 times at 20 seconds each and then centrifuged at 7354*g* for 5 min. The A_{280} of the resulting supernatant was measured using the NanoDrop Lite (Thermo) to measure protein concentration. The supernatants were then diluted to a concentration of 2 mg/mL of total protein to a final volume of 20 μL. 4 μL of 6X SDS loading dye was added to the 20 μL of the supernatant and the samples were then heated for 95°C for 15 min and then 20 μL of each sample was loaded onto the SDS-PAGE gel. The gel was run at 200V for 50 min at 4°C and then stained with Coomassie Blue.

3.6 References

1 Pfleger, B.F., Gossing, M. & Nielsen, J. Metabolic engineering strategies for microbial synthesis of oleochemicals. *Metab Eng* **29**, 1-11 (2015).

2 Schweizer, E. & Hofmann, J. Microbial type I fatty acid synthases (FAS): Major players in a network of cellular FAS systems. *Microbiol Mol Biol Rev* **68**, 501-517 (2004).

3 Dellomonaco, C., Clomburg, J.M., Miller, E.N. & Gonzalez, R. Engineered reversal of the beta-oxidation cycle for the synthesis of fuels and chemicals. *Nature* **476**, 355-U131 (2011).

4 McMahon, M.D. & Prather, K.L. Functional screening and in vitro analysis reveal thioesterases with enhanced substrate specificity profiles that improve short-chain fatty acid production in Escherichia coli. *Appl Environ Microbiol* **80**, 1042-50 (2014).

5 Zheng, Y. et al. Boosting the free fatty acid synthesis of Escherichia coli by expression of a cytosolic Acinetobacter baylyi thioesterase. *Biotechnol Biofuels* **5**, 76 (2012).

6 Jing, F. et al. Phylogenetic and experimental characterization of an acyl-ACP thioesterase family reveals significant diversity in enzymatic specificity and activity. *BMC Biochem* **12**, 44 (2011).

7 Pollard, M.R., Anderson, L., Fan, C., Hawkins, D.J. & Davies, H.M. A Specific Acyl-Acp Thioesterase Implicated in Medium-Chain Fatty-Acid Production in Immature Cotyledons of Umbellularia-Californica. *Arch Biochem Biophys* **284**, 306-312 (1991).

8 Leber, C. & Da Silva, N.A. Engineering of Saccharomyces cerevisiae for the synthesis of short chain fatty acids. *Biotechnol Bioeng* **111**, 347-58 (2014).

9 Choi, Y.J. & Lee, S.Y. Microbial production of short-chain alkanes. *Nature* **502**, 571-574 (2013).

10 Kovacic, F. et al. Structural and Functional Characterisation of TesA - A Novel Lysophospholipase A from Pseudomonas aeruginosa. *PLoS One* **8** (2013).

11 Grisewood, M.J. et al. Computational Redesign of Acyl-ACP Thioesterase with Improved Selectivity toward Medium-Chain-Length Fatty Acids. *Acs Catalysis* **7**, 3837-3849 (2017).

12 Zornetzer, G.A., Tanem, J., Fox, B.G. & Markley, J.L. The Length of the Bound Fatty Acid Influences the Dynamics of the Acyl Carrier Protein and the Stability of the Thioester Bond. *Biochemistry* **49**, 470-477 (2010).

13 Worthington, A.S. et al. Probing the compatibility of type II ketosynthase-carrier protein partners. *Chembiochem* **9**, 2096-2103 (2008).

14 Mukherjee, K., Bhattacharyya, S. & Peralta-Yahya, P. GPCR-Based Chemical Biosensors for Medium-Chain Fatty Acids. *ACS Synth Biol* **4**, 1261-1269 (2015).

15 Voelker, T.A. & Davies, H.M. Alteration of the Specificity and Regulation of Fatty-Acid Synthesis of Escherichia-Coli by Expression of a Plant Medium-Chain Acyl-Acyl Carrier Protein Thioesterase. *J Bacteriol* **176**, 7320-7327 (1994).

16 Bond-Watts, B.B., Bellerose, R.J. & Chang, M.C.Y. Enzyme mechanism as a kinetic control element for designing synthetic biofuel pathways. *Nat Chem Biol* **7**, 222-227 (2011).

17 Steen, E.J. et al. Microbial production of fatty-acid-derived fuels and chemicals from plant biomass. *Nature* **463**, 559-U182 (2010).

18 Kozakov, D. et al. How good is automated protein docking? *Proteins-Structure Function and Bioinformatics* **81**, 2159-2166 (2013).

19 Kelley, L.A., Mezulis, S., Yates, C.M., Wass, M.N. & Sternberg, M.J.E. The Phyre2 web portal for protein modeling, prediction and analysis. *Nat Protocols* **10**, 845-858 (2015).

20 Chan, D.I. & Vogel, H.J. Current understanding of fatty acid biosynthesis and the acyl carrier protein. *Biochem J* **430**, 1-19 (2010).

21 Zhang, Y.M., Wu, B., Zheng, J. & Rock, C.O. Key residues responsible for acyl carrier protein and beta-ketoacyl-acyl carrier protein reductase (FabG) interaction. *J Biol Chem* **278**, 52935-43 (2003).

22 Zhang, Y.M. et al. Identification and analysis of the acyl carrier protein (ACP) docking site on beta-ketoacyl-ACP synthase III. *J Biol Chem* **276**, 8231-8 (2001).

23 Rafi, S. et al. Structure of acyl carrier protein bound to FabI, the FASII enoyl reductase from Escherichia coli. *J Biol Chem* **281**, 39285-39293 (2006).

24 Finzel, K. et al. Probing the Substrate Specificity and Protein-Protein Interactions of the E. coli Fatty Acid Dehydratase, FabA. *Chem Biol* **22**, 1453-1460 (2015).

25 Ford, T.J. & Way, J.C. Enhancement of E-coli acyl-CoA synthetase FadD activity on medium chain fatty acids. *PeerJ* **3** (2015).

26 Shiue, E. & Prather, K.L.J. Improving D-glucaric acid production from myo-inositol in E. coli by increasing MIOX stability and myo-inositol transport. *Metab Eng* **22**, 22-31 (2014).

27 Yu, G. et al. Enzymatic functions of wild tomato methylketone synthases 1 and 2. *Plant Physiol* **154**, 67-77 (2010).

28 Torella, J.P. et al. Tailored fatty acid synthesis via dynamic control of fatty acid elongation. *Proc Natl Acad Sci U S A* **110**, 11290-5 (2013).

CHAPTER 4. PHYLOGENY-BASED ENGINEERING OF A BACTERIAL ACYL-ACP THIOESTERASE

4.1 Abstract

Medium-chain fatty acids (MCFAs) are useful intermediates to polymer precursors such as medium-chain (MC)-dicarboxylic acids and fuels such as MC-alkyl esters. *De novo* production of MCFAs in *E. coli* can be initiated by bacterial acyl-ACP thioesterases (TEs). However, bacterial acyl-ACP TEs have broad specificity. Here, we engineered *A. baylyi* thioesterase (AbTE) using a phylogeny-based approach. Previous work suggests sequence variability in acyl-ACP thioesterases corresponds to the differences in specificity observed in the phylogeny. We hypothesized that mutations of variable positions on AbTE to those residues found in the phylogeny could improve AbTE specificity/activity towards MCFAs. Alignment of AbTE related sequences from the *Acinetobacter* genus identified variable positions at the thioesterase surface. The phylogeny-based single mutant screen revealed two surface residue mutations that improved MCFA titers ~1.3-fold over wild-type. Combination of the phylogeny-derived AbTE mutations did not improve MCFA titers over the single mutants, but led to a shift in fatty acid profile toward long-chain fatty acids. The phylogeny-based mutagenesis was marginally successful in improving MCFA titers.

4.2 Introduction

Microbial production of MCFAs with carbon lengths between C8-C12 has garnered interest because of their use as intermediates for the synthesis of industrially important

chemicals. MCFAs can be derivatized into MC-alkanes that can be used as direct fuels, MC dicarboxylic acids that can be used as nylon precursors or MC α-olefins that can be used as precursors to bioplastics. Production of MCFAs in microbes has been primarily achieved via metabolic engineering coupled to the expression of medium-chain specific TEs. During fatty acid biosynthesis, the growing fatty acyl intermediate is bound to the acyl-carrier protein (ACP) and will elongate to a C16-C18 carbon chain[1]. However, the fatty acyl-ACP intermediates can be prematurely terminated into a C8-C12 free fatty acid by medium-chain specific acyl-ACP TEs.

Medium-chain TEs from plant (FatBs) or bacterial origin have been expressed in *E. coli* for *de novo* production of MCFAs[2-4]. Plant medium-chain acyl-ACP TEs (FatBs) have generally narrower product profile than their bacterial counterparts, such as the *Umbellularia californica* TE cleaving primarily C12:0 chain lengths[5], and *Cuphea palustris* TE preferring C8:0[6]. Aside from the well characterized *Escherichia coli* TesA, bacterial acyl-ACP thioesterases have been less explored than plant TEs. Bacterial acyl-ACP TEs are typically non-specific, thus can cleave multiple chain lengths of acyl-ACP intermediates in addition to MC-acyl-ACPs[7]. However, they may have the advantage of compatibility with the *E. coli* ACP. The importance of homologous ACP/TE pairs has been shown with algal ACPs which preferentially interact with the algal TE versus a plant TE[8]. Two groups have engineered TEs for the improved production of MCFAs. In both cases, the long-chain TE from *E. coli*, TesA, has been engineered and used for improved MCFA production in *E. coli*[9, 10]. Recently, we have engineered AbTE for improved surface interaction with *E. coli* ACP resulting in an improvement in MCFA titers (Chapter 3).

The bacterial acyl-ACP TE from *Acinetobacter baylyi* ADP1 (AbTE) has resulted in high MCFA titers when expressed in *E. coli*[3] and *Saccharomyces cerevisiae*[11]. *A. baylyi* ADP1 contains other thioesterases in addition to AbTE which vary in chain-length specificity, preferring long-chain fatty acyl-ACP substrates C14-C16[12]. Additionally, the *Acinetobacter* genus is a source of other lipid tailoring enzymes, such as wax ester synthases[13] and acyl-CoA reductases[14]. We hypothesized that there are other acyl-ACP thioesterases within the *Acinetobacter* genus that may have improved activity or substrate specificity that could result in higher MCFAs titers when expressed in *E. coli*. Rather than synthesizing every acyl-ACP thioesterases like-enzyme in the *Acinetobacter* genus and testing it for MCFA production in *E. coli*, we used phylogeny to identify highly variable positions in acyl-ACP thioesterases like-enzymes and the residues at these positions. These variable positions in *Acinetobacter* TEs may indicate if and where specificity for different chain-lengths deviated in evolution since these TEs come from a common ancestor. With this in mind, we introduced the residues from the phylogeny on to the AbTE scaffold and tested them for MCFA production.

Bacterial genera such as *Bacteroides*, *Lactobacillus*, and *Clostridium* contain acyl-ACP TEs within their respective genus with varying chain-length specificities[7]. Therefore, mutating one or several residues in a bacterial TE sequence to residues found in related TEs from the same genus may be enough to alter specificity. This approach was seen in plant acyl-ACP TEs where mutation of a variable surface methionine on the *Arabidopsis thaliana* FatB TE to a threonine found on *A. thaliana* FatA TE resulted in a profile shift from primarily C14:0 and C16:1 to exclusively C16:1 when expressed in *E. coli*[15]. Thus, mutating variable residues that mimic evolutionary changes within a genus

can improve desired enzyme characteristics and are also less likely of decreasing enzyme activity versus highly conserved residues[16].

Phylogeny-based mutagenesis has also been shown to improve other enzyme characteristics, such as kinetics, thermostability, and expression. Analysis of CYP450 monooxygenase sequences identified a variable residue[17] that was mutated in a *Bacilllus megaterium* CYP450 and led to either shifts in chemo-selectivity or improved turnover (~10-fold)[18]. In another example, a phylogenetic tree and alignment was used to engineer ancestral residues in a *Thermus thermophilus* 3-isopropylmalate dehydrogenase which led to an improvement in thermostability by 3.2˚C[19]. Finally, a dual family shuffling and phylogeny-based mutagenesis approach led to improved expression of a mammalian serum paraoxonases in *E. coli*[20].

Here, we investigate the insertion of residues on AbTE at variable positions found in acyl-ACP TE-like enzymes from the *Acinetobacter* genus and their effect on MCFA production in *E. coli*. Analogous to the aforementioned TEs in other bacterial genera, the TE-like enzymes from the *Acinetobacter* genus may have enough diversity in specificity to assist in altering the specificity of AbTE. We observe that both the two best single mutants and combination of these mutations led to ~1.3-fold improvement in MCFA titers versus wild-type. Analysis of the effect of these mutations on MCFA profile and overall MCFA titers will give a better understanding of AbTE structure-function relationships and identify key positions that can further mutated or be translatable to other bacterial acyl-ACP TEs.

4.3 Results

4.3.1 Multiple sequence alignment of A. baylyi sequence

The Protein Basic Local Alignment Search Tool (BLAST) was used to search for sequences related to AbTE from 7 species of *Acinetobacter*. Specifically, the alignment contained 100 sequences annotated with different descriptors such as arylesterases, multifunctional acyl-CoA thioesterase I, GDSL-like proteins, and even hypothetical proteins within the genus.

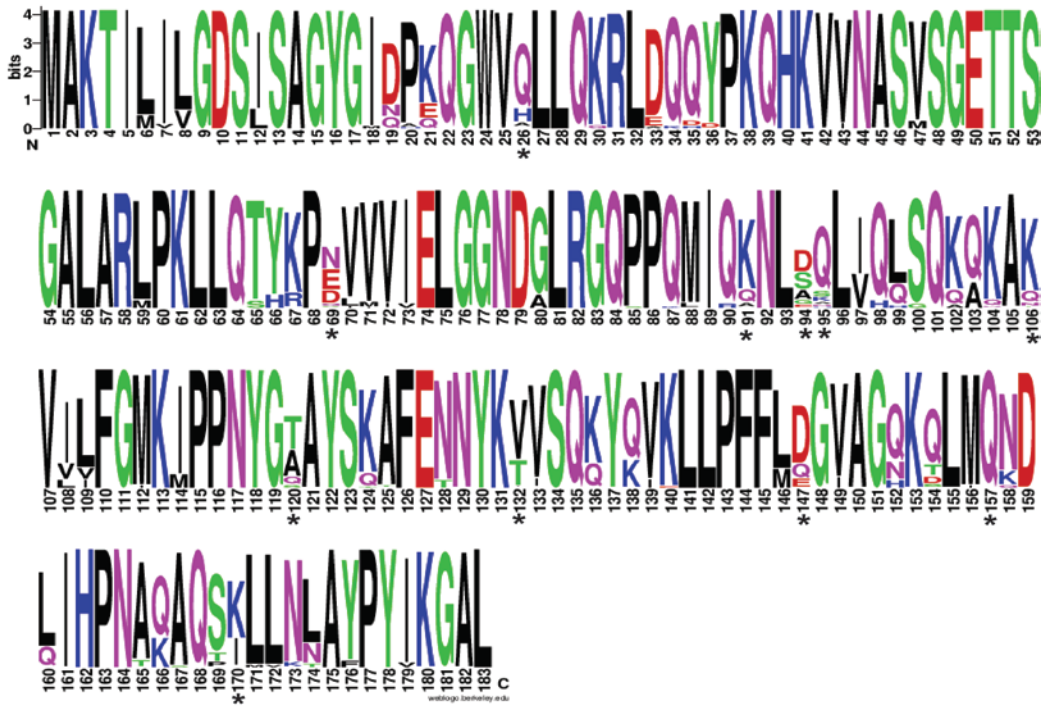

Figure 4.1: AbTE multiple sequence alignment, Logo representation. Positions with only one letter are conserved while those with multiple letters are variable. Letter size corresponds to the most dominant residue in that position. Positions targeted for mutation contain an asterisk. Sequence logo created using http://weblogo.berkeley.edu/logo.cgi.

The multiple sequence alignment of acyl-ACP TE-like enzymes showed highly conserved and variable positions. The residues of the catalytic triad (S11, D159, H162) and the oxyanion hole (G48, N77)[3] of aligned TEs are highly conserved (Figure 4.1). The

alignment also shows stretches of highly conserved regions (e.g. SVSGETTS and ELGGND), which are found in loops surrounding the active site. To quantify variability of the sequence alignment at each position, we used the Protein Variability Server (PVS)[21] as this server has been previously used to determine the sequence variability in enzymes such as a squalene monooxygenases[22] and glutamate racemases[23]. Within the server, we analyzed sequence variability using the Wu-Kabat (WK) variability coefficient, which has been historically been used for identification of variable regions within antibodies[24.] The WK method is a well-established method of describing a residue's susceptibility to evolutionary change and can emphasize regions of variation[21]. Through the WK analysis, any position with a WK value >1 is considered variable as higher values correspond with higher variability. To reduce the number AbTE mutants to screen, we prioritized positions with WK values of ≥5 (Table 4-1), corresponding to 11 variable positions with a total of 44 single mutations (Figure 4.1). The selected variable residues were generally present on the TE surface and away from the binding pocket region (Figure 4.2), consistent with the notion that surface residues are more prone to evolutionary changes than core residues[25].

Table 4-1: *A. baylyi* thioesterase amino acids selected for mutagenesis

Selected AAs with values >5	***other AAs in phylogeny***	***WK value***
N69	Q, D, S, E, H	16.5
E94	G, A, S, D, V	16.054
T120	A, Q, S	8.609
A26	H, Q, T, L, N	7.92
K95	A, Q, S, N, R	7.714
I170	K, Q, E, L	7.615

Table 4-1: Continued

Selected AAs with values >5	***other AAs in phylogeny***	***WK value***
S154	T, D, Q, H	7.071
S91	K, Q, A	6.429
K106	Q, T, H, N	6.266
V132	I, T, L	5.91
D147	E, Q, N	5.739

4.3.2 In vivo activity screen of phylogeny-based AbTE mutants

Expression of AbTE variants carrying mutations at variable positions in *E. coli* resulted in different levels of MCFA titers when compared to wild-type AbTE, which produced about ~27.2 mg/L of MCFAs (Figure 4.2). The vast majority of the AbTE mutations resulted in similar or decreased secreted MCFA titers when compared to wild-type levels. All mutations at positions A26 resulted in decreased MCFA titers with AbTE:A26H producing ~6.4 mg/L of MCFAs, the largest decrease at position 26. Residue E94 was mutated to hydrophobic, charged, and polar residues, with AbTE:E94S resulting in the only improvement reaching ~35.8 mg/L of MCFA titers (*P*-value = 0.01), a ~1.3-fold improvement versus wild-type. In contrast, AbTE:K106H resulted in an almost complete reduction of MCFA titers, reaching only ~0.97 mg/L which was attributed only to C12s. Residue T120 was previously hypothesized to interact with *E. coli* ACP (Chapter 3). Although the mutant AbTE:T120R in Chapter 3 did not improve MCFA titers, AbTE:T120A had a statistically significant improvement versus wild-type (*P*-value = 0.04) resulting in ~35.9 mg/L (~1.3-fold improvement versus wild-type).

ACP interactions could be enhanced with AbTE:T120A as this mutant may have established hydrophobic interactions with a matching residue on the ACP (e.g. M or V). Lastly, a complete loss of MCFA production was observed AbTE:I170L, which was interesting given the strong similarity between I and L.

From the single mutant screen, only two mutations, AbTE:E94S and AbTE:T120A led to a ~1.3-fold improvement in MCFA titers over wild-type both producing ~35 mg/L. In contrast, the largest decreases in MCFA titers were seen with AbTE:K106H which produced ~0.97 mg/L, and AbTE:I170L which produced zero. Interestingly, AbTE:T120A and AbTE:E94S resulted in no major alteration of chain-length specificity compared to the wild-type profile (Table 2-1).

Figure 4.2: Phylogeny-based engineering of *Acinetobacter baylyi* thioesterase (AbTE). A. AbTE homology model with docked C12 acyl-phosphopantetheine (green) and variable residues targeted for mutagenesis. B. Secreted fatty acid titers of *E .coli* expressing AbTE single mutants. Green represents medium-chain fatty acids (MCFAs, C8-C12) and purple represents long-chain fatty acids (LCFAs, C14-C18). Asterisk corresponds to mutants with MCFA titer titers with statistically significant improvement, *P*-value < 0.05 (two-tailed *t* test). All experiments were done in triplicate and the error bars represent the standard deviation from the mean.

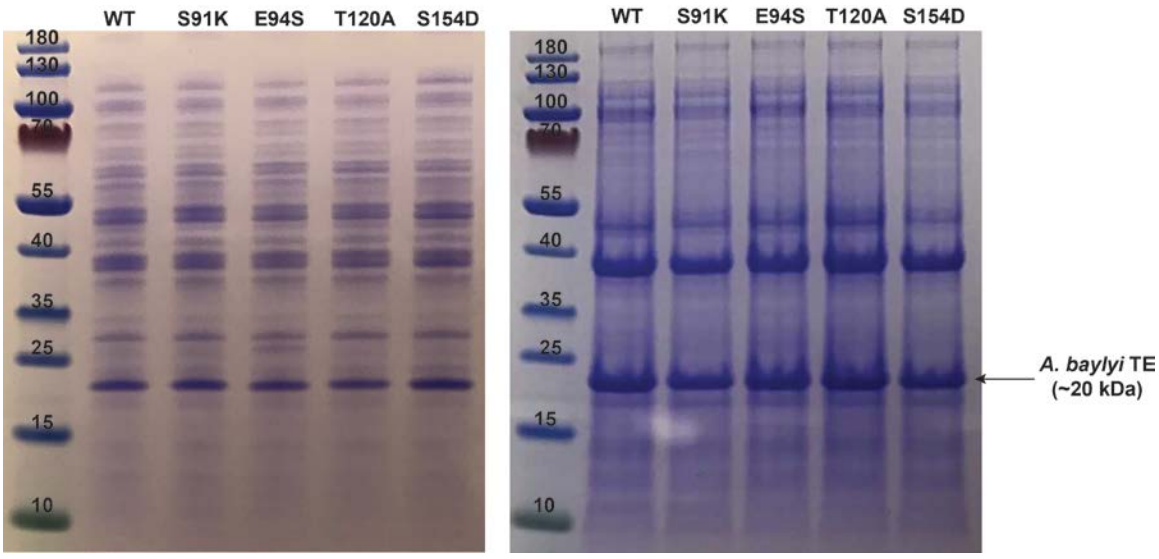

Figure 4.3: Coomassie stained SDS-PAGE gel of induced *E. coli* cultures expressing AbTE:WT and best variants from single mutant screen- AbTE:S91K, AbTE:E94S, AbTE:T120A, and AbTE:S154D. Left gel is the soluble total protein, right gel is insoluble total protein.

Because the two mutations that improved MCFA titers are present on AbTE's surface (Figure 4.4), we hypothesized that the improved titers may be the result of enhanced expression of AbTE in *E. coli*. We ran an SDS-PAGE gel from *E. coli* expressing AbTE wild-type, the single mutants leading to the highest MCFA titers (AbTE:E94S and AbTE:T120A), and two single mutants that introduced charged residues on the AbTE (AbTE:S91K and AbTE:S154D), which could also improve solubility. The SDS-PAGE gel shows comparable soluble and insoluble expression levels (~20kDa) for all AbTE enzymes (Figure 4.3). Improved soluble expression is not likely the reason for improved MCFA titers in the case of AbTE:E94S and AbTE:T120A. The improvements in MCFA titers may be a result of enhanced TE turnover, which could be proven in the future with an *in vitro* assay that tests acyl-ACP intermediates of all chain-lengths.

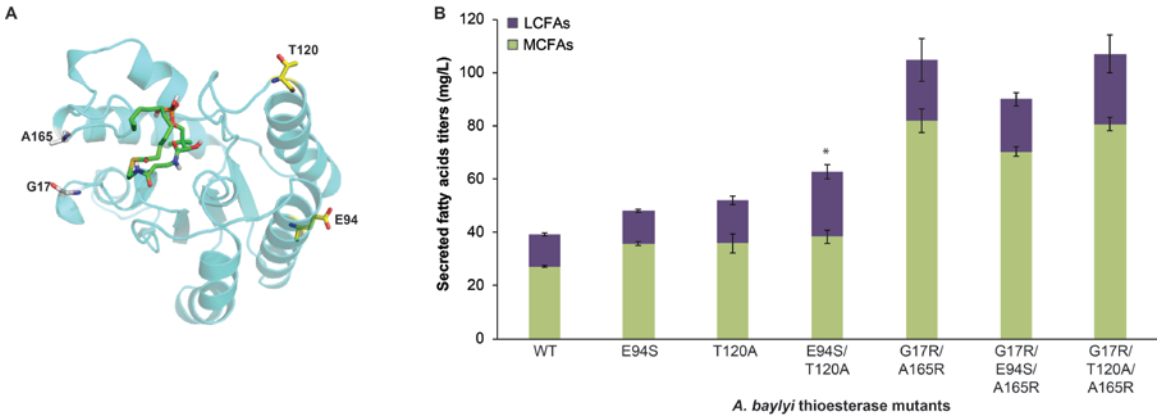

Figure 4.4: Combination of phylogeny-based *A. baylyi* TE mutants. A. AbTE homology model with docked C12 acyl-phosphopantetheine (green). Residues shown to improve titers from single mutant screen are shown in yellow and previously mutated interface residues (white). B. Secreted fatty acid levels of *E .coli* expressing best AbTE single mutants and mutant combinations. Green represents medium-chain fatty acids (MCFAs, C8-C12) and purple represents long-chain fatty acids (LCFAs, C14-C18). Asterisk corresponds to mutant with MCFA titers with statistically significant improvement, *P*-value < 0.05 (two-tailed *t* test). All experiments were done in triplicate and the error bars represent the standard deviation from the mean.

Next, we next determined if combining the two single mutations lead to improved MCFA titers. The AbTE:E94S/T120A mutation did not improve MCFA titers over the single mutants but retained a ~1.3-fold improvement of MCFAs over wild-type (*P*-value < 0.05) (Figure 4.4). Unexpectedly, AbTE:E94S/T120A displayed a ~3.6-fold improvement in C14 titers; from ~2.9 mg/L in wild-type to ~10.9 mg/L (*P*-value < 0.05), demonstrating that mutating both residues had impact on specificity.

Table 4-2: Chain-length profile of best MCFA-producing *A. baylyi* TE phylogeny-based mutants.

	% of Total Fatty Acids						
A. baylyi TE variant	***C8***	***C10***	***C12***	***C14***	***C16***	***C18***	***MCFAs***
WT	27.65	16.41	25.32	7.60	9.67	13.36	69.37

Table 4-2: Continued

	% of Total Fatty Acids						
***A. baylyi* TE variant**	***C8***	***C10***	***C12***	***C14***	***C16***	***C18***	***MCFAs***
E94S	30.46	18.20	25.84	7.58	7.81	10.10	74.51
T120A	29.18	16.76	23.36	12.21	9.19	9.29	69.31
E94S/T120A	25.34	13.88	21.98	17.24	12.09	9.48	61.20
G17R/A165R	39.03	19.64	19.63	10.90	6.21	4.58	78.30
G17R/E94S/A165R	40.30	19.76	18.12	9.82	6.75	5.25	78.18
G17R/T120A/A165R	36.92	19.08	19.41	13.09	7.08	4.42	75.41

Next, we combined the improved phylogeny-based mutations in AbTE with our previous best mutant, AbTE:G17R/A165R (Chapter 3). The triple mutant AbTE:G17R/E94S/A165R produced ~70.3 mg/L while AbTE:G17R/T120A/A165R produced ~80.7 mg/L of MCFAs. Neither of the mutants produced higher MCFA titers versus AbTE:G17R/A165R that produced ~82 mg/L (Figure 4.4).

4.4 Discussion

We hypothesized that mutating the highly variable positions of AbTE to residues found in TE sequences of *Acinetobacter* origin would improve AbTE specificity/activity towards MCFAs. In the single mutant screen, we identified two mutants that marginally improved MCFAs versus wild-type. Both AbTE:E94S and AbTE:T120A resulted in a similar ~1.3-fold improvement in MCFA titers producing ~35 mg/L. The combination of these two mutants, AbTE:E94S/T120A, led to ~38 mg/L but there was no statistically significant improvement over the single mutants. However, a change of specificity was observed for AbTE:E94S/T120A where C14 fatty acids were improved 3.6-fold versus wild-type, showing the mutant pair had a synergistic effect on specificity. Attempting to

improve the previously developed thioesterase mutant (AbTE:G17R/A165R, Chapter 3) by the addition of E94S or T120A, did not result have an additive effect on MCFA titers.

As a secondary analysis, we also explored the total fatty acid titers (C8-C18) generated from our mutant screen. From the same single mutant screen, five different mutants, AbTE:E94G, E94S, T120Q, S154H and S154D caused minor, but statistically significant (*P*-value < 0.05) improvements (~1.2-fold) for total fatty acid titers versus wild-type (Figure 4.5). The best producer, AbTE:E94S, reached ~48 mg/L of total fatty acids which was also the best MCFA producer.

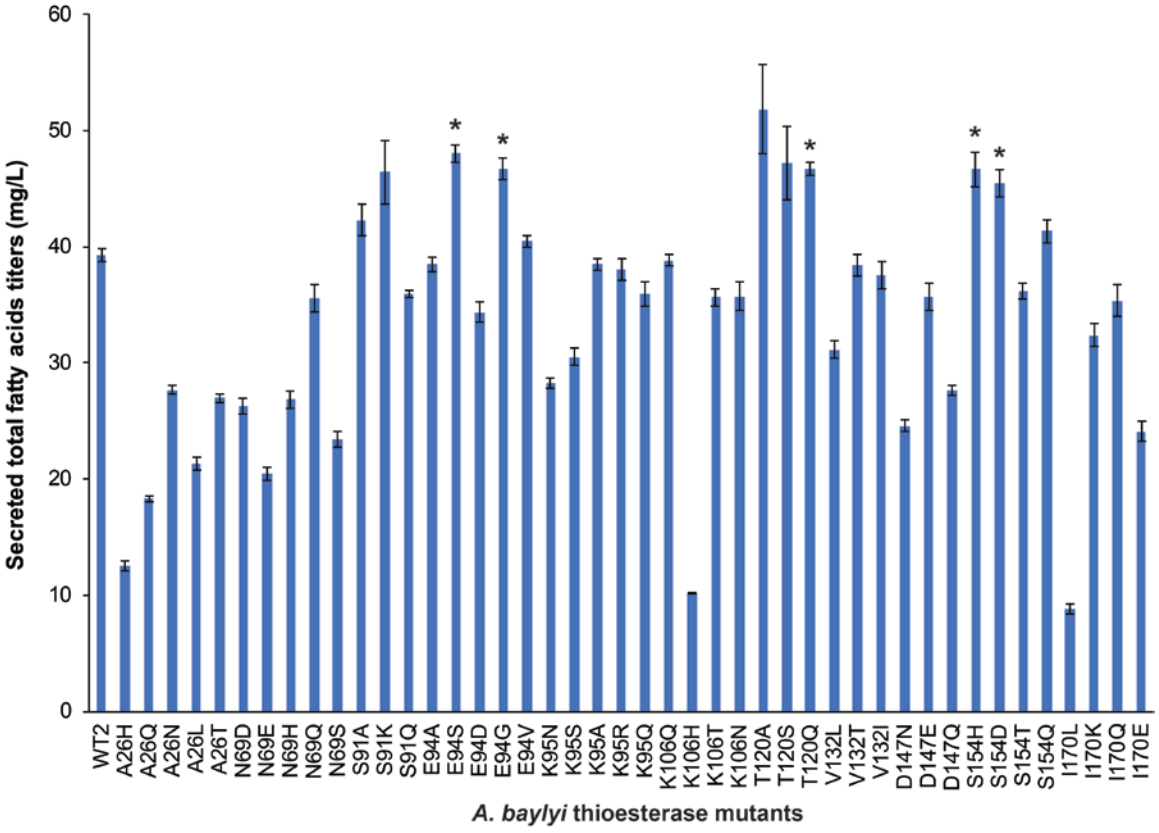

Figure 4.5: Secreted total fatty acid (C8-C18) titers of *E .coli* expressing AbTE single mutants. . Asterisk corresponds to mutant with fatty acid titers with statistically significant improvement, *P*-value < 0.05 (two-tailed *t* test). All

experiments were done in triplicate and the error bars represent the standard deviation from the mean.

Mutations of highly variable residues in AbTE were generally unsuccessful in engineering improved medium-chain fatty acid producing AbTE variants. Second, the improvements we did observe did not result in drastic changes in specificity towards MCFA production. However, the change in specificity toward C14 fatty acids that was observed in AbTE:E94S/T120A may be indicative of a widening of the binding pocket because of the substitution of E and T to smaller residues. In the future, we can further mutate these positions with different combinations of bulkier residues that may have the opposite effect and fine tune specificity towards medium-chain lengths.

Additionally, the AbTE residues that were selected for mutation were the most variable residues in the alignment, but exploring variable residues strictly in the binding pocket region could be more ideal for altering specificity. In the future, we will target variable residues in the binding pocket region and mutate them to those in the phylogeny and to bulky residues that may shift the broad specificity to a narrower profile. Further, instead of mutating highly variable residues, we could target positions with intermediate variability which might be more influential as evolutionary changes in specificity-influencing residues might be subtler.

4.5 Materials and Methods

4.5.1 Reagents

Chemical standards were purchased from: C8:0 (A10991, Alfa Aesar); C10:0 (A15658, Alfa Aesar); C12:0 (A12492, Alfa Aesar); C14:0 (A10257, Alfa Aesar); C16:0 (P0006, TCI); C18:0 (S0080, TCI).

4.5.2 Multiple sequence alignment and variability analysis

The amino acid sequence of AbTE (removed periplasmic sequence) was inputted into the Protein Basic Local Alignment Search Tool (BLAST) via the non-redundant protein sequences database. The resulting top 100 best aligned sequences that had 86-83% identity relative to AbTE were obtained and truncated at the same position as AbTE. The sequences were then aligned through ClustalX2. The alignment was then submitted to Protein Variability server for amino acid variability analysis. Positions with a Wu Kabat (WK) value of ≥5 were chosen for mutagenesis.

4.5.3 AbTE mutant generation

Site-directed mutagenesis was performed using the QuikChange protocol with some modifications. PCR reaction: 0.8 ng/μL of template, 2.5 ng/μL of each primer, 1X iProof HF polymerase buffer, 0.02 U/μL iProof polymerase (BioRad), 0.5 mM dNTPs to 50 μL final volume. Details on templates and primers (degenerate codons were used in some cases to generate multiple mutants at once) used are in APPENDIX D. Construction of Site directed mutants). Thermocycler protocol: 95°C 1min, 17 cycles: [95°C 50sec, 60°C 50sec, 72°C 2min 30sec], 72°C 7 min. DpnI (0.4U/ μL) was added to PCR reaction and incubated at 37°C for 1 hour and heat inactivated at 65°C for 20 min. 10 μL of reaction was transformed into competent DH10B *E. coli* (PPY252) cells.

4.5.4 Fatty acid production and quantification

MCFA Production: Overnight cultures of PPY251 (*E. coli* MG1655) expressing AbTE:WT or AbTE variants were diluted 1:50 in 5 mL of M9 media (0.5% glucose, amp^{100}) and grown at 37°C, 250 r.p.m. until reaching an OD_{600} = 0.3-0.4. The cells were then induced with 500 μM of IPTG (500 mM stock) and grown at 30°C, 250 r.p.m. for 24 hrs. *Fatty acid analysis:* To analyze secreted fatty acids, *E. coli* cultures were vortexed for 3 sec, 600 μL of culture removed and centrifuged for 10 min at 7354*g*. Next, 400 μL of the supernatant was removed for derivatization. *Fatty acid* d*erivatization:* Fatty acids were derivatized to fatty acid methyl esters and analyzed via GC-MS as described in Chapter 3. To the 400 μL of sample, 50 μL of 10% (wt/vol) NaCl, 50 μL of glacial acetic acid, 20 μL of 90.5 mg/L nonanoic acid (internal standard), and 200 μL of ethyl acetate were added and the mixture was vortexed for 5 sec. The mixture was then centrifuged at 12,098*g* for 10 min. Methyl esters were generated by mixing 100 μL of the ethyl acetate layer with 900 μL of a 30:1 mixture of methanol and 37% (vol/vol) HCl in a 2 mL microcentrifuge tube, vortexed for 5 sec, and incubated at 50°C for 1 hr. After cooling to room temperature, 500 μL of water and 500 μL of hexanes were added. The mixture was vortexed for 5 sec, 100μL of the hexane layer was taken and mixed with 400 μL of ethyl acetate for analysis via GC-MS. *FAME quantification*: The samples were analyzed using Agilent 7890A/Agilent 5975 MS detector using a DB-5MS column. The inlet temperature was set to 300°C, flow at 1 mL/min, the oven at 70°C for 1 min, ramp at 30°C/min to 290°C, and held for 1 min at 290°C. Standard curves of C8-C18 fully saturated FAMEs (Alfa Aesar/TCI) were used for sample quantification.

4.5.5 A. baylyi thioesterase homology model and docking

Phyre2[26] software was used to generate the AbTE homology model. The model was generated in intensive mode and the confidence in the model for 181 residues (99%) was modelled at >90% accuracy. C12-acyl- phosphopantetheine was docked onto the homology model via AutoDock Vina[27] to show observe binding pocket region.

4.5.6 SDS-PAGE gel

PPY1331, PPY1567, PPY1570, PPY1581, and PPY1589 were grown overnight in LB media amp^{100}. The overnight cultures were diluted 1:50 in 1 mL of M9 media (0.5% glucose, amp^{100}). The cultures were grown at 37^{o}C, 250 rpm. until they reached an OD_{600} = 0.30-0.40. The cells were then induced with 500 μM of IPTG (500 mM stock conc.) and grown at 30^{o}C, 250 rpm for 24 hours. 1 mL of sample was removed from the culture medium and centrifuged 5 min at 7354*g*. The supernatants were discarded, and the pellet resuspended in 200 μL of 1X PBS. The cell suspension was sonicated 2 times at 20 seconds each and then centrifuged at 7354*g* for 5 min. The A_{280} of the resulting supernatant was measured using the NanoDrop Lite (Thermo) to measure protein concentration. For the soluble fractions, the supernatants were diluted to a concentration of 2 mg/mL of total protein to a final volume of 20 μL. 4 μL of 6X SDS loading dye was added to the 20 μL of the supernatant. For the insoluble fraction, 40 μL of 6X SDS loading dye was added. Samples were then heated for 95°C for 15 min and then 20 μL of each sample of the soluble was loaded onto the SDS-PAGE gel while 10 μL was loaded for insoluble fraction samples. The gel was run at 200V for 50 min at 4°C and then stained with Coomassie Blue.

4.6 References

1 Janssen, H.J. & Steinbuchel, A. Fatty acid synthesis in Escherichia coli and its applications towards the production of fatty acid based biofuels. Biotechnol Biofuels 7, 7 (2014).

2 Lennen, R.M., Braden, D.J., West, R.A., Dumesic, J.A. & Pfleger, B.F. A process for microbial hydrocarbon synthesis: Overproduction of fatty acids in Escherichia coli and catalytic conversion to alkanes. *Biotechnol Bioeng* **106**, 193-202 (2010).

3 Zheng, Y. et al. Boosting the free fatty acid synthesis of Escherichia coli by expression of a cytosolic Acinetobacter baylyi thioesterase. *Biotechnol Biofuels* **5**, 76 (2012).

4 Torella, J.P. et al. Tailored fatty acid synthesis via dynamic control of fatty acid elongation. *Proc Natl Acad Sci U S A* **110**, 11290-5 (2013).

5 Voelker, T.A. & Davies, H.M. Alteration of the specificity and regulation of fatty acid synthesis of Escherichia coli by expression of a plant medium-chain acyl-acyl carrier protein thioesterase. *J Bacteriol* **176**, 7320-7 (1994).

6 Dehesh, K., Edwards, P., Hayes, T., Cranmer, A.M. & Fillatti, J. Two novel thioesterases are key determinants of the bimodal distribution of acyl chain length of Cuphea palustris seed oil. *Plant Physiol* **110**, 203-10 (1996).

7 Jing, F.Y. et al. Phylogenetic and experimental characterization of an acyl-ACP thioesterase family reveals significant diversity in enzymatic specificity and activity. *Bmc Biochem* **12** (2011).

8 Beld, J., Blatti, J.L., Behnke, C., Mendez, M. & Burkart, M.D. Evolution of acyl-ACP-thioesterases and beta-ketoacyl-ACP-synthases revealed by protein-protein interactions. *J Appl Phycol* **26**, 1619-1629 (2014).

9 Choi, Y.J. & Lee, S.Y. Microbial production of short-chain alkanes. *Nature* **502**, 571-574 (2013).

10 Grisewood, M.J. et al. Computational Redesign of Acyl-ACP Thioesterase with Improved Selectivity toward Medium-Chain-Length Fatty Acids. *Acs Catalysis* **7**, 3837-3849 (2017).

11 Zhu, Z. et al. Expanding the product portfolio of fungal type I fatty acid synthases. *Nat Chem Biol* **13**, 360-362 (2017).

12 Ukey, R., Holmes, W.E., Bajpai, R. & Chistoserdov, A.Y. Evaluation of thioesterases from Acinetobacter baylyi for production of free fatty acids. *Can J Microbiol* **63**, 321-329 (2017).

13 Stoveken, T., Kalscheuer, R., Malkus, U., Reichelt, R. & Steinbuchel, A. The wax ester synthase/acyl coenzyme A:diacylglycerol acyltransferase from Acinetobacter sp.

strain ADP1: characterization of a novel type of acyltransferase. *J Bacteriol* **187**, 1369-76 (2005).

14 Ishige, T. et al. Wax ester production from n-alkanes by Acinetobacter sp. strain M-1: ultrastructure of cellular inclusions and role of acyl coenzyme A reductase. *Appl Environ Microbiol* **68**, 1192-5 (2002).

15 Mayer, K.M. & Shanklin, J. Identification of amino acid residues involved in substrate specificity of plant acyl-ACP thioesterases using a bioinformatics-guided approach. *Bmc Plant Biology* **7** (2007).

16 Bloom, J.D. & Arnold, F.H. In the light of directed evolution: pathways of adaptive protein evolution. *Proc Natl Acad Sci U S A* **106 Suppl 1**, 9995-10000 (2009).

17 Seifert, A. & Pleiss, J. Identification of selectivity-determining residues in cytochrome P450 monooxygenases: a systematic analysis of the substrate recognition site 5. *Proteins* **74**, 1028-35 (2009).

18 Seifert, A. et al. Rational design of a minimal and highly enriched CYP102A1 mutant library with improved regio-, stereo- and chemoselectivity. *Chembiochem* **10**, 853-61 (2009).

19 Watanabe, K., Ohkuri, T., Yokobori, S. & Yamagishi, A. Designing thermostable proteins: ancestral mutants of 3-isopropylmalate dehydrogenase designed by using a phylogenetic tree. *J Mol Biol* **355**, 664-74 (2006).

20 Khersonsky, O. et al. Directed evolution of serum paraoxonase PON3 by family shuffling and ancestor/consensus mutagenesis, and its biochemical characterization. *Biochemistry* **48**, 6644-54 (2009).

21 Garcia-Boronat, M., Diez-Rivero, C.M., Reinherz, E.L. & Reche, P.A. PVS: a web server for protein sequence variability analysis tuned to facilitate conserved epitope discovery. *Nucleic Acids Res* **36**, W35-W41 (2008).

22 Chua, N.K., Howe, V., Jatana, N., Thukral, L. & Brown, A.J. A Conserved Degron Containing an Amphipathic Helix Regulates the Cholesterol-Mediated Turnover of Human Squalene Monooxygenase, a Rate-Limiting Enzyme in Cholesterol Synthesis. *J Biol Chem* (2017).

23 Kaushik, P., Jain, C.K., Gabrani, R. & Singh, T.R. Study on variability assessment and evolutionary relationships of glutamate racemase in Pseudomonas species. *Interdiscip Sci* **5**, 247-57 (2013).

24 Johnson, G. & Wu, T.T. Kabat Database and its applications: 30 years after the first variability plot. *Nucleic Acids Res* **28**, 214-218 (2000).

25 Jackson, E.L., Ollikainen, N., Covert, A.W., 3rd, Kortemme, T. & Wilke, C.O. Amino-acid site variability among natural and designed proteins. *PeerJ* **1**, e211 (2013).

26 Kelley, L.A. & Sternberg, M.J. Protein structure prediction on the Web: a case study using the Phyre server. *Nat Protoc* **4**, 363-71 (2009).

27 Trott, O. & Olson, A.J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem* **31**, 455-61 (2010).

CHAPTER 5. FUTURE OUTLOOK

5.1 Conclusions and Future outlook

5.1.1 Improving microbial titers of pinene to be cost-competitive

Pinene production titers reached ~32 mg/L by fusing the geraniol diphosphate synthase and pinene synthase from *A. grandis*. At these levels (~1.2% of theoretical yield) we would need to improve ~26-fold in order to be cost competitive with JP-10 fuel. To achieve this, the initial route to improving titers would be to optimize several parameters such as expression fine-tuning via combinatorial screening of plasmid with different copy numbers and promoters. Second, the major bottleneck enzyme, the pinene synthase, would need to be engineered for improved activity. Recently, a *P. taeda* pinene synthase variant (Y346H/Q457L) was engineered via a high-throughput screen for improved activity in *E. coli* resulting in ~140 mg/L[1]. From our screen, the pinene synthase from *A. grandis* resulted in higher titers than the one from *P. taeda*. Thus, we could adapt the beneficial mutations into the *A. grandis* pinene synthase that also has Y346 and Q457 while using the available colorimetric screen[1] to improve activity further. Lastly, the pinene producing pathway may benefit from a transfer to another chassis such as *S. cerevisiae*. Yeast has recently been shown to be robust chassis for monoterpene production via engineering for geraniol production, reaching a titer of 1.69 g/L[2]. An

optimized pinene construct and enzyme system in *E. coli* could be then be shuttled to yeast to boost titers.

5.1.2 High-throughput screening of medium-chain fatty acid producer strains

In Chapter 3 and 4 there were two successful rational engineering attempts at improving medium-chain fatty acid titers. While the thioesterase's *in vivo* activity has been improved several fold, it can be further optimized for activity and specificity as our best mutants still cleave multiple fatty acyl chain lengths. As there is no crystal structure of the *A. baylyi* thioesterase, the most viable option is to screen mutants for improved microbially produced MCFAs via the GPCR-based sensor used in Chapter 2. The sensor was able to discriminate varying MCFA levels of *E. coli* producers expressing thioesterase variants. However, to apply the sensor to high-throughput screening of MCFA-producing microbes, we must improve the sensor's dynamic range to better distinguish between low, intermediate and high MCFA producers. One avenue to improve the dynamic range is through engineering the GPCR, OR1G1 for binding to a specific fatty acid chain length (e.g. C10). The olfactory GPCR used in the sensor interacts with C8-C12 fatty acids and derivatives such as aldehydes and alcohols with varying affinities, while preferring C10 chain lengths[3]. As a result, our sensor has the highest fold-change for C10 fatty acids[4]. However, mixtures of multiple odorant chemicals can have antagonistic effects on OR1G1 resulting in dampened signal[3]. Thus, the mixture of fatty acids and other metabolites in a producer cell's supernatant may be causing a negative effect on sensor signal. A more specific GPCR variant would remove the competition between fatty acid chain-lengths and possibly other metabolites in the binding pocket of OR1G1 hence improving dynamic range. Once the sensor can easily

distinguish high MCFA producers from low producers, we can screen larger libraries of MCFA producer cells that express *A. baylyi* thioesterase variants.

Aside from engineering the terminal thioesterase, another path for improving MCFA titers can be via engineering of upstream fatty acid biosynthesis enzymes that generate medium-chain acyl-ACP/CoA substrates, which are limiting in MCFA production. As mentioned in Chapter 1, upstream chain-length determining enzymes (i.e. ketosynthases) have been rationally engineered in several microbes for MCFA production with varying success. Therefore, we could use our sensor to screen for MCFAs secreted from *E. coli* producer cells that contain libraries of ketosynthases and thereby guide engineering of improved medium-chain acyl-ACP formation.

5.2 References

1 Tashiro, M. et al. Bacterial production of pinene by a laboratory-evolved pinene synthase. *ACS Synth Biol* **5**, 1011–1020 (2016).

2 Zhao, J.Z. et al. Dynamic control of ERG20 expression combined with minimized endogenous downstream metabolism contributes to the improvement of geraniol production in Saccharomyces cerevisiae. *Microb Cell Fact.* **16** (2017).

3 Sanz, G., Schlegel, C., Pernollet, J.C. and Briand, L. Comparison of odorant specificity of two human olfactory receptors from different phylogenetic classes and evidence for antagonism. *Chem Senses* **30**, 69-80 (2005).

4 Mukherjee, K., Bhattacharyya, S. and Peralta-Yahya, P. GPCR-Based Chemical Biosensors for Medium-Chain Fatty Acids. *ACS Synth Biol* **4**, 1261-1269 (2015).

APPENDIX A. PLASMID TABLES

Table A-1: Chapter 1 plasmids

Strain#	Plasmid name	Description	Reference
JBEI-3085	pBbA5c-MevT-MBI	p15A, Cm^r, PLacUV5, atoB-HMGS-HMGR-MK-PMK-PMD-idi.	Bokinsky, et al. 2011 *PNAS*
JBEI-2318	pSynAgGPPS	Commercially synthesized GPPS from *Abies grandis* (Genescript)	This study, Chapter 1
JBEI-4261	pSynPaGPPS	Commercially synthesized GPPS from *Picea abies* (Genescript)	This study, Chapter 1
JBEI-4260	pSynPtGPPS	Commercially synthesized GPPS from *Pinus taeda* (Genescript)	This study, Chapter 1
JBEI-3169	pSynAgPS	Commercially synthesized PS from *Abies grandis* (Genescript)	This study, Chapter 1
JBEI-4262	pSynPaPS	Commercially synthesized PS from *Picea abies* (Genescript)	This study, Chapter 1
JBEI 3195	pSynPtPS	Commercially synthesized PS from *Pinus taeda* (Genescript)	This study, Chapter 1
JBEI-4888	pAgGPPS	pMB1, Amp^r, P_{Trc}, *A. grandis* GPPS, signal peptide removed	This study, Chapter 1
JBEI-4892	pPaGPPS	pMB1, Amp^r, P_{Trc}, *P. abies GP*PS signal peptide removed	This study, Chapter 1
JBEI-4887	pPtGPPS	pMB1, Amp^r, P_{Trc}, *P. taeda* GPPS signal peptide removed	This study, Chapter 1

Table A-1: Continued

Strain	Plasmid name	Description	Reference
JBEI-3176	ptAgPS	pMB1, Amp^r, P_{Trc}, *A. grandis* PS signal peptide removed (truncated at $RR(X)_8W$)	This study, Chapter 1
JBEI-667	ptPaPS	pMB1, Amp^r, P_{Trc}, *P. abies* PS signal peptide removed (truncated at $RR(X)_8W$)	This study, Chapter 1
JBEI-3197	ptPtPS	pMB1, Amp^r, P_{Trc}, *P. taeda* PS signal peptide removed (truncated at $RR(X)_8W$)	This study, Chapter 1
PPY33	pLead	Commercially synthesized leader sequence for *Pinus taeda* and *Picea abies* (Operon)	This study, Chapter 1
PPY151	pAgPS	pMB1, Amp^r, P_{Trc}, *A. grandis* PS signal peptide removed (truncated at RRGKSI)	This study, Chapter 1
PPY153	pPaPS	pMB1, Amp^r, P_{Trc}, *P. abies* PS signal peptide removed (truncated at RRGKSA)	This study, Chapter 1
PPY152	pPtPS	pMB1, Amp^r, P_{Trc}, *P. taeda* PS signal peptide removed (truncated at RRTLPG)	This study, Chapter 1
PPY325	pHis6-PaPS	pMB1, Amp^r, P_{Trc}, His_6-tag at the N-terminus of *P. abies* PS signal peptide removed (truncated at RRGKSA)	This study, Chapter 1
PPY324	pHis6-PtPS	pMB1, Amp^r, P_{Trc}, His_6-tag at the N-terminus of *P. taeda* PS signal peptide removed (truncated at RRTLPG)	This study, Chapter 1
PPY272	$pHis_6$-AgPS	pMB1, Amp^r, P_{Trc}, His_6-tag at the N-terminus of *A. grandis* PS signal peptide removed (truncated at RRGKSI)	This study, Chapter 1

Table A-1: Continued

Strain	Plasmid name	Description	Reference
PPY304	pHis6-PtGPPS	pMB1, Amp^r, P_{Trc}, His_6-tag at the N-terminus of *P. taeda* GPPS signal peptide removed	This study, Chapter 1
PPY274	pHis6-PaGPPS	pMB1, Amp^r, P_{Trc}, His_6-tag at the N-terminus of *P. abies* GPPS signal peptide removed	This study, Chapter 1
PPY273	pHis6-AgGPPS	pMB1, Amp^r, P_{Trc}, His_6-tag at the N-terminus of *A. grandis* GPPS signal peptide removed	This study, Chapter 1
PPY175	pAgPS-AgGPPS	pMB1, Amp^r, P_{Trc}, *A. grandis* PS- *A. grandis* GPPS	This study, Chapter 1
PPY179	pAgPS-PaGPPS	pMB1, Amp^r, P_{Trc}, *A. grandis* PS- *P. abies* GPPS	This study, Chapter 1
PPY180	pAgPS-PtGPPS	pMB1, Amp^r, P_{Trc}, *A. grandis* PS- *P. taeda* GPPS	This study, Chapter 1
PPY176	pPaPS-AgGPPS	pMB1, Amp^r, P_{Trc}, *P. abies* PS- *A. grandis* GPPS	This study, Chapter 1
PPY178	pPaPS-PaGPPS	pMB1, Amp^r, P_{Trc}, *P. abies* PS- *P. abies* GPPS	This study, Chapter 1
PPY177	pPaPS-PtGPPS	pMB1, Amp^r, P_{Trc}, *P. abies* PS- *P. taeda* GPPS	This study, Chapter 1
PPY128	pPtPS-AgGPPS	pMB1, Amp^r, P_{Trc}, *P. taeda* PS- *A. grandis* GPPS	This study, Chapter 1
PPY127	pPtPS-PaGPPS	pMB1, Amp^r, P_{Trc}, *P. taeda* PS- *P. abies* GPPS	This study, Chapter 1

Table A-1: Continued

Strain	Plasmids	Description	Reference
PPY126	pPtPS-PtGPPS	pMB1, Amp^r, P_{Trc}, *P. taeda* PS- *P. taeda* GPPS	This study, Chapter 1
PPY316	pAgGPPS-(GSG)-AgPS	pMB1, Amp^r, P_{Trc}, *A. grandis* GPPS-(GSG)-*A. grandis* PS	This study, Chapter 1
PPY320	pAgGPPS-$(GSG)_3$-AgPS	pMB1, Amp^r, P_{Trc}, *A. grandis* GPPS-$(GSG)_3$-*A. grandis* PS	This study, Chapter 1
JBEI-3933	pAgGPPS-$(GSG)_2$-AgPS	pMB1, Amp^r, P_{Trc}, *A. grandis* GPPS-$(GSG)_2$-*A. grandis* PS	This study, Chapter 1
PPY404	pAgPS-$(GSG)_2$-AgGPPS	pMB1, Amp^r, P_{Trc}, *A. grandis* PS-$(GSG)_2$-*A. grandis* GPPS	This study, Chapter 1
PPY128	pAgGPPS-$(GSG)_2$-PaPS	pMB1, Amp^r, P_{Trc}, *A. grandis* GPPS-$(GSG)_2$-*P. abies* PS	This study, Chapter 1
PPY216	pAgGPPS-$(GSG)_2$-PtPS	pMB1, Amp^r, P_{Trc}, *A. grandis* GPPS-$(GSG)_2$-*P. taeda* PS	This study, Chapter 1
PPY69	pPaGPPS-$(GSG)_2$-AgPS	pMB1, Amp^r, P_{Trc}, *P. abies* GPPS-$(GSG)_2$-*A. grandis* PS	This study, Chapter 1
PPY129	pPaGPPS-$(GSG)_2$-PaPS	pMB1, Amp^r, P_{Trc}, *P. abies* GPPS-$(GSG)_2$-*P. abies* PS	This study, Chapter 1
PPY215	pPaGPPS-$(GSG)_2$-PtPS	pMB1, Amp^r, P_{Trc}, *P. abies* GPPS-$(GSG)_2$-*P. taeda* PS	This study, Chapter 1
PPY68	pPtGPPS-$(GSG)_2$-AgPS	pMB1, Amp^r, P_{Trc}, *P. taeda* GPPS-$(GSG)_2$-*A. grandis* PS	This study, Chapter 1
PPY127	pPtGPPS-$(GSG)_2$-PaPS	pMB1, Amp^r, P_{Trc}, *P. taeda* GPPS-$(GSG)_2$-*P. abies* PS	This study, Chapter 1

Table A-1: Continued

PPY125	pPtGPPS-$(GSG)_2$-PtPS	pMB1, Amp^r, P_{Trc}, *P. taeda* GPPS-$(GSG)_2$-*P. taeda* PS	This study, Chapter 1

Table A-2: Chapter 3 plasmids

Strain #	**Plasmid name**	**Description**	**Reference**
PPY269	pKM260	pESC-His3-P_{TEF1}-OR1G1- P_{ADH1}	Mukherjee, et al. 2015 *ACS Synth. Biol.*
PPY586	pKM586	pRS415-Leu2-P_{FIG1}-GFP	Mukherjee, et al. 2015 *ACS Synth. Biol.*
PPY1023	pSS185	pMB1-Amp^r-P_{TRC}-AgGPPS-$(GSG)_2$-AgPS	Sarria, et al. 2014 *ACS Synth. Biol.*, Chapter 1
PPY1090	pSS174	pMB1-Amp^r-P_{TRC}-CnTE	This study, Chapter 3
PPY1148	pSS183	pMB1-Amp^r-P_{TRC}-CpTE	This study, Chapter 3
PPY1236	pSS192	pMB1-Amp^r-P_{TRC}-AbTE:WT	This study, Chapter 3
PPY1237	pSS193	pMB1-Amp^r-P_{TRC}-UcTE	This study, Chapter 3
PPY1310	pSS196	pMB1-Amp^r-P_{TRC}-AbTE:G17R	This study, Chapter 3
PPY1311	pSS197	pMB1-Amp^r-P_{TRC}-AbTE:T120R	This study, Chapter 3
PPY1312	pSS198	pMB1-Amp^r-P_{TRC}-AbTE:A121R	This study, Chapter 3

Table A-2: Continued

Strain #	Plasmid name	Description	Reference
PPY1320	pSS199	pMB1-Amp^r-P$_{TRC}$-AbTE:A165R	This study, Chapter 3
PPY1321	pSS200	pMB1-Amp^r-P$_{TRC}$-AbTE:G17R/A165R	This study, Chapter 3
PPY1322	pSS201	pMB1-Amp^r-P$_{TRC}$-AbTE:A121R/A165R	This study, Chapter 3
PPY1323	pSS202	pMB1-Amp^r-P$_{TRC}$-AbTE:T120R/A165R	This study, Chapter 3
PPY1326	pSS203	pMB1-Amp^r-P$_{TRC}$-AbTE:G17R/A121R	This study, Chapter 3
PPY1327	pSS204	pMB1-Amp^r-P$_{TRC}$-AbTE:G17R/T120R	This study, Chapter 3
PPY1328	pSS205	pMB1-Amp^r-P$_{TRC}$-AbTE:T120R/A121R	This study, Chapter 3
PPY1329	pSS206	pMB1-Amp^r-P$_{TRC}$-AbTE:S11A	This study, Chapter 3
PPY1393	pSS208	pMB1-Amp^r-P$_{TRC}$-AbTE:G17R/A165R/T120R	This study, Chapter 3
PPY1394	pSS209	pMB1-Amp^r-P$_{TRC}$-AbTE:G17R/A165R/A121R	This study, Chapter 3
PPY1395	pSS210	pMB1-Amp^r-P$_{TRC}$-AbTE:G17R/A165R/N158D	This study, Chapter 3
PPY1499	pSS251	pMB1-Amp^r-P$_{TRC}$-AbTE:G17E	This study, Chapter 3
PPY1500	pSS252	pMB1-Amp^r-P$_{TRC}$-AbTE:G17E/A165E	This study, Chapter 3
PPY1501	pSS253	pMB1-Amp^r-P$_{TRC}$-AbTE:G17E/A165R	This study, Chapter 3
PPY1502	pSS254	pMB1-Amp^r-P$_{TRC}$-AbTE:G17R/A165E	This study, Chapter 3

Table A-3: Chapter 4 plasmids

Strain #	Plasmid name	Description	Reference
PPY1236	pSS192	pMB1-Amp^r-P$_{TRC}$-AbTE:WT	Chapter 3
PPY1321	pSS200	pMB1-Amp^r-P$_{TRC}$-AbTE:G17R/A165R	Chapter 3
PPY1462	pSS214	pMB1-Amp^r-P$_{TRC}$-AbTE:A26H	This study, Chapter 4
PPY1463	pSS215	pMB1-Amp^r-P$_{TRC}$-AbTE:A26Q	This study, Chapter 4
PPY1464	pSS216	pMB1-Amp^r-P$_{TRC}$-AbTE:A26N	This study, Chapter 4
PPY1465	pSS217	pMB1-Amp^r-P$_{TRC}$-AbTE:A26T	This study, Chapter 4
PPY1466	pSS218	pMB1-Amp^r-P$_{TRC}$-AbTE:N69D	This study, Chapter 4
PPY1467	pSS219	pMB1-Amp^r-P$_{TRC}$-AbTE:N69E	This study, Chapter 4
PPY1468	pSS220	pMB1-Amp^r-P$_{TRC}$-AbTE:N69H	This study, Chapter 4
PPY1469	pSS221	pMB1-Amp^r-P$_{TRC}$-AbTE:N69Q	This study, Chapter 4
PPY1470	pSS222	pMB1-Amp^r-P$_{TRC}$-AbTE:N69S	This study, Chapter 4
PPY1471	pSS223	pMB1-Amp^r-P$_{TRC}$-AbTE:S91K	This study, Chapter 4
PPY1472	pSS224	pMB1-Amp^r-P$_{TRC}$-AbTE:S91Q	This study, Chapter 4
PPY1473	pSS225	pMB1-Amp^r-P$_{TRC}$-AbTE:E94A	This study, Chapter 4
PPY1474	pSS226	pMB1-Amp^r-P$_{TRC}$-AbTE:E94S	This study, Chapter 4
PPY1475	pSS227	pMB1-Amp^r-P$_{TRC}$-AbTE:E94D	This study, Chapter 4
PPY1476	pSS228	pMB1-Amp^r-P$_{TRC}$-AbTE:E94G	This study, Chapter 4

Table A-3: Continued

Strain #	Plasmid name	Description	Reference
PPY1477	pSS229	pMB1-Amp^r-P$_{TRC}$-AbTE:E94V	This study, Chapter 4
PPY1478	pSS230	pMB1-Amp^r-P$_{TRC}$-AbTE:K95N	This study, Chapter 4
PPY1479	pSS231	pMB1-Amp^r-P$_{TRC}$-AbTE:K95S	This study, Chapter 4
PPY1480	pSS232	pMB1-Amp^r-P$_{TRC}$-AbTE:K95A	This study, Chapter 4
PPY1481	pSS233	pMB1-Amp^r-P$_{TRC}$-AbTE:K95Q	This study, Chapter 4
PPY1482	pSS234	pMB1-Amp^r-P$_{TRC}$-AbTE:K106Q	This study, Chapter 4
PPY1483	pSS235	pMB1-Amp^r-P$_{TRC}$-AbTE:K106H	This study, Chapter 4
PPY1484	pSS236	pMB1-Amp^r-P$_{TRC}$-AbTE:K106T	This study, Chapter 4
PPY1485	pSS237	pMB1-Amp^r-P$_{TRC}$-AbTE:T120A	This study, Chapter 4
PPY1486	pSS238	pMB1-Amp^r-P$_{TRC}$-AbTE:T120S	This study, Chapter 4
PPY1487	pSS239	pMB1-Amp^r-P$_{TRC}$-AbTE:T120Q	This study, Chapter 4
PPY1488	pSS240	pMB1-Amp^r-P$_{TRC}$-AbTE:V132T	This study, Chapter 4
PPY1489	pSS241	pMB1-Amp^r-P$_{TRC}$-AbTE:V132I	This study, Chapter 4
PPY1490	pSS242	pMB1-Amp^r-P$_{TRC}$-AbTE:D147E	This study, Chapter 4
PPY1491	pSS243	pMB1-Amp^r-P$_{TRC}$-AbTE:D147Q	This study, Chapter 4
PPY1492	pSS244	pMB1-Amp^r-P$_{TRC}$-AbTE:S154H	This study, Chapter 4

Table A-3: Continued

Strain #	Plasmid name	Description	Reference
PPY1493	pSS245	pMB1-Amp^r-P$_{TRC}$-AbTE:S154D	This study, Chapter 4
PPY1494	pSS246	pMB1-Amp^r-P$_{TRC}$-AbTE:S154T	This study, Chapter 4
PPY1495	pSS247	pMB1-Amp^r-P$_{TRC}$-AbTE:S154Q	This study, Chapter 4
PPY1496	pSS248	pMB1-Amp^r-P$_{TRC}$-AbTE:I170K	This study, Chapter 4
PPY1497	pSS249	pMB1-Amp^r-P$_{TRC}$-AbTE:I170Q	This study, Chapter 4
PPY1498	pSS250	pMB1-Amp^r-P$_{TRC}$-AbTE:I170E	This study, Chapter 4
PPY1523	pSS258	pMB1-Amp^r-P$_{TRC}$-AbTE:A26L	This study, Chapter 4
PPY1524	pSS259	pMB1-Amp^r-P$_{TRC}$-AbTE:S91A	This study, Chapter 4
PPY1525	pSS260	pMB1-Amp^r-P$_{TRC}$-AbTE:K95R	This study, Chapter 4
PPY1526	pSS261	pMB1-Amp^r-P$_{TRC}$-AbTE:K106N	This study, Chapter 4
PPY1527	pSS262	pMB1-Amp^r-P$_{TRC}$-AbTE:V132L	This study, Chapter 4
PPY1528	pSS263	pMB1-Amp^r-P$_{TRC}$-AbTE:D147N	This study, Chapter 4
PPY1529	pSS264	pMB1-Amp^r-P$_{TRC}$-AbTE:I170L	This study, Chapter 4
PPY1550	pSS270	pMB1-Amp^r-P$_{TRC}$-AbTE:E94S/T120A	This study, Chapter 4
PPY1554	pSS274	pMB1-Amp^r-P$_{TRC}$-AbTE:G17R/E94S/A165R	This study, Chapter 4
PPY1555	pSS275	pMB1-Amp^r-P$_{TRC}$-AbTE:G17R/T120A/A165R	This study, Chapter 4

APPENDIX B. STRAIN TABLES

Table B-1: Chapter 1 strains

Strain #	Description	Reference
PPY251	*Escherichia coli* strain MG1655 (pine production experiments)	ATCC 47076
PPY252	*Escherichia coli* strain DH10B (cloning)	Invitrogen

Plasmids from Chapter 1 (Table A-1: Chapter 1 plasmids) were transformed into PPY251 (MG1655) for pinene production experiments or PPY252 (DH10B) for cloning purposes. There are no PPY251 strains transformed with these plasmids on strain collection.

Table B-2: Chapter 3 strains

Strain #	Description	Reference
PPY11	W303: MATa, leu2−3, trp1−1, can1−100, ura3−1, ade2−1, his3−11	ATCC 208352
PPY140	PPY11 Δ*far1*, Δ*sst2*, Δ*ste2*	Mukherjee, et al. 2015 *ACS Synth. Biol.*
PPY643	PPY140, pKM260 & pKM586	Mukherjee, et al. 2015 *ACS Synth. Biol.*
PPY252	DH10B	Invitrogen
PPY251	MG1655	ATCC 47076
PPY260	DH5α	Invitrogen
PPY1151	BW25113 Δ*fadE*739::kan	Keio collection
PPY1236	PPY252, pSS192	This study, Chapter 3
PPY1331	PPY251, pSS192	This study, Chapter 3
PPY1332	PPY251, pSS206	This study, Chapter 3

Table B-2: Continued

PPY1333	PPY251, pSS196	This study, Chapter 3
PPY1334	PPY251, pSS199	This study, Chapter 3
PPY1335	PPY251, pSS198	This study, Chapter 3
PPY1336	PPY251, pSS197	This study, Chapter 3
PPY1337	PPY251, pSS203	This study, Chapter 3
PPY1338	PPY251, pSS204	This study, Chapter 3
PPY1339	PPY251, pSS205	This study, Chapter 3
PPY1340	PPY251, pSS200	This study, Chapter 3
PPY1341	PPY251, pSS201	This study, Chapter 3
PPY1342	PPY251, pSS202	This study, Chapter 3
PPY1396	PPY251, pSS208	This study, Chapter 3
PPY1397	PPY251, pSS209	This study, Chapter 3
PPY1398	PPY251, pSS210	This study, Chapter 3
PPY1403	PPY251, pSS174	This study, Chapter 3
PPY1404	PPY251, pSS183	This study, Chapter 3
PPY1405	PPY251, pSS193	This study, Chapter 3
PPY1406	BL21	This study, Chapter 3
PPY1407	PPY260, pSS192	This study, Chapter 3
PPY1408	PPY1152, pSS192	This study, Chapter 3
PPY1409	PPY1461, pSS192	This study, Chapter 3
PPY1503	PPY251, pSS251	This study, Chapter 3
PPY1504	PPY251, pSS252	This study, Chapter 3
PPY1505	PPY251, pSS253	This study, Chapter 3
PPY1506	PPY251, pSS254	This study, Chapter 3

Table B-3: Chapter 4 strains

Strain #	Description	Reference
PPY252	DH10B	Invitrogen
PPY251	MG1655	ATCC 47076
PPY1331	PPY251, pSS192	Sarria, 2017
PPY1340	PPY251, pSS200	Sarria, 2017
PPY1558	PPY251, pSS214	This study, Chapter 4
PPY1559	PPY251, pSS215	This study, Chapter 4
PPY1560	PPY251, pSS216	This study, Chapter 4
PPY1561	PPY251, pSS217	This study, Chapter 4
PPY1562	PPY251, pSS218	This study, Chapter 4
PPY1563	PPY251, pSS219	This study, Chapter 4
PPY1564	PPY251, pSS220	This study, Chapter 4
PPY1565	PPY251, pSS221	This study, Chapter 4
PPY1566	PPY251, pSS222	This study, Chapter 4
PPY1567	PPY251, pSS223	This study, Chapter 4
PPY1568	PPY251, pSS224	This study, Chapter 4
PPY1569	PPY251, pSS225	This study, Chapter 4
PPY1570	PPY251, pSS226	This study, Chapter 4
PPY1571	PPY251, pSS227	This study, Chapter 4
PPY1572	PPY251, pSS228	This study, Chapter 4
PPY1573	PPY251, pSS229	This study, Chapter 4
PPY1574	PPY251, pSS230	This study, Chapter 4
PPY1575	PPY251, pSS231	This study, Chapter 4
PPY1576	PPY251, pSS232	This study, Chapter 4
PPY1577	PPY251, pSS233	This study, Chapter 4
PPY1578	PPY251, pSS234	This study, Chapter 4
PPY1579	PPY251, pSS235	This study, Chapter 4
PPY1580	PPY251, pSS236	This study, Chapter 4
PPY1581	PPY251, pSS237	This study, Chapter 4
PPY1582	PPY251, pSS238	This study, Chapter 4

Table B-3: Continued

PPY1583	PPY251, pSS239	This study, Chapter 4
PPY1584	PPY251, pSS240	This study, Chapter 4
PPY1585	PPY251, pSS241	This study, Chapter 4
PPY1586	PPY251, pSS242	This study, Chapter 4
PPY1587	PPY251, pSS243	This study, Chapter 4
PPY1588	PPY251, pSS244	This study, Chapter 4
PPY1589	PPY251, pSS245	This study, Chapter 4
PPY1590	PPY251, pSS246	This study, Chapter 4
PPY1591	PPY251, pSS247	This study, Chapter 4
PPY1592	PPY251, pSS248	This study, Chapter 4
PPY1593	PPY251, pSS249	This study, Chapter 4
PPY1594	PPY251, pSS250	This study, Chapter 4
PPY1595	PPY251, pSS258	This study, Chapter 4
PPY1596	PPY251, pSS259	This study, Chapter 4
PPY1597	PPY251, pSS260	This study, Chapter 4
PPY1598	PPY251, pSS261	This study, Chapter 4
PPY1599	PPY251, pSS262	This study, Chapter 4
PPY1600	PPY251, pSS263	This study, Chapter 4
PPY1601	PPY251, pSS264	This study, Chapter 4
PPY1605	PPY251, pSS270	This study, Chapter 4
PPY1609	PPY251, pSS274	This study, Chapter 4
PPY1610	PPY251, pSS275	This study, Chapter 4

APPENDIX C. PRIMER TABLES

Table C-1: Chapter 1 primers

Name	Sequence
SS31	TCGACCGTTCTTTCGACGTTGGTTCTGGTGGTTCTGGT
SS32	TGGTGACCCGCGATACGA
SS33	GCGCGTCGTATCGCGG
SS34	TCTCATCCGCCAAAACAGCCTTACAGCGGAACGGTTTCA
SS35	ACATCGCGTCTCGTCAGAACGGCTCCGGTGGCTCC
SS36	TGGAAGTCACCCATACGA
SS37	CGTCGTATGGGTGACTTC
SS38	TCTCATCCGCCAAAACAGCCTTACAGGGTAACCGGGTC
SS45	ACATCGCGTCTCGTCAGAACGGTTCTGGTGGTTCTGGT
SS46	TCTCATCCGCCAAAACAGCCTTACAGCGGAACAGATTCC
SS49	TCGACCGTTCTTTCGACGTTGGTTCTGGTGGTTCTGGT
SS50	TCTCATCCGCCAAAACAGCCTTACAGCGGAACAGATTCC
SS54	TCTCATCCGCCAAAACAGCCTTACAGCGGAACGGTTTCA
SS55	TCGACCGTTCTTTCGACGTTGGTTCTGGTGGTTCTGGTGCGCGTC GTACCCTGTTC
SS61	ACGTGGCATTTCGTCAGAATGGCTCCGGTGGCTCCG
SS62	TCTCATCCGCCAAAACAGCCTTACAGGGTAACCGGGTC
SS63	TCGACCGTTCTTTCGACGTTGGCTCCGGTGGCTCCG
SS64	ACGTGGCATTTCGTCAGAATGGTTCTGGTGGTTCTGGT
SS65	TCTCATCCGCCAAAACAGCCTTACAGCGGAACGGTTTCAA
SS66	ATCGACCGTTCTTTCGACGTTGGTTCTGGTGGTTCTGGT
SS67	TCTCATCCGCCAAAACAGCCTTACAGCGGAACGGTTTCAA
SS68	ACGTGGCATTTCGTCAGAATGGTTCTGGTGGTTCTGGTGCGCGT CGTACCCTGTT
SS69	ACATCGCGTCTCGTCAGAACGGTTCTGGTGGTTCTGGTGCGCGT CGTACCCTGTT

Table C-1: Continued

SS70	CACACAGGAAACAGACCATGCATCACCATCACCATCACGAATTTGACTTCAACAAATACA
SS71	CACACAGGAAACAGACCATGCATCACCATCACCATCACGGTTTCGACCTGCGTCG
SS72	CACACAGGAAACAGACCATGCATCACCATCACCATCACGAGTTCGACTTCGACAAATA
SS74	CACACAGGAAACAGACCATGCATCACCATCACCATCACGCGCGTCGTACCCTGTT
SS75	CACACAGGAAACAGACCATGCATCACCATCACCATCACCGTCGTGGTAAATCTGCG
SS76	TAGAGGATCCCCGGG
SS77	CCCGGGGATCCTCTAG
SS78	CATGGTCTGTTTCCTGTGT
SS79	CACACAGGAAACAGACCATGCATCACCATCACCATCACCTGGTCGTCGTGGTAAATCT
SS87	TCTCATCCGCCAAAACAGCCTTACAGCGGAACAGATTCC
SS89	ACGTGGCATTTCGTCAGAATGGTTCTGGTGGTTCTGGTGGTTCTGGT
SS90	ACGTGGCATTTCGTCAGAATGGTTCTGGTGCGCGTCGTGGTAAATCT
SS117	TGCTGGAATCTGTTCCGCTGGGTTCTGGTGGTTCTGGTGAATTTGACTTCAACAAATACATG
SS118	TCTCATCCGCCAAAACAGCCTTAATTCTGACGAAATGCC
PPY58	TCCGCTGTAACCCGGGGATCATTGTGAGCGGATAACAATT
PPY59	TCTCATCCGCCAAAACAG
PPY60	TACCCTGTAACCCGGGGATCATTGTGAGCGGATAACAATT
PPY61	TCCGCTGTAACCCGGGGATCATTGTGAGCGGATAACAATT
PPY90	GGTTCTGGCAAATATTCTG
PPY91	ACCGCTTCTGCGTTCTG
PPY608	CCCGGGTTAATTCTGACGAAATGCCACG
PPY630	CCATGGAATTTGACTTCAACAAATAC
PPY642	CCATGGCGCGTCGTATCGCGGGTCACC

Table C-1: Continued

PPY643	CCCGGGTTACAGCGGAACGGTTTCA
PPY782	CCCGGGTTACAGCGGAACAGATTCC
PPY969	CCATGGCGCGTCGTATGGGTGACTTC
PPY1210	CCATGGAGTTCGACTTCGACA
PPY1219	CCATGGGTTTCGACCTGCGT
PPY1212	CCCGGGTTACAGGGTAACCGGGTCGA
PPY1220	CCATGGCGCGTCGTATGGGTGACTTC
PPY1225	CCCGGGTTAAACGTCGAAAGAACGGTC
PPY1227	CCCGGGTTAGTTCTGACGAGACGCGAT
BW33	GATAGATTTACCACGACGACCAGAACCACCAGAACC
BW34	CGTCGTGGTAAATCTATC

Table C-2: Chapter 3 primers

Name	**Sequence (SDM primers have codon change in bold)**
SS455	TAACAATTTCACACAGGAAACAGACCATGGATGTTGCCAGATTGGTCTATG
SS456	CCTGCAGGTCGACTCTAGAGGATCCCCGGGTTATTTAGATTCAGTTGGATGCAAACCC
SS541	CGACAGTCTGAGTGCGGGTTATAGAATTAACCCCGAACAGGGCTGG
SS542	CCAGCCCTGTTCGGGGTTAATTCTATAACCCGCACTCAGACTGTCG
SS543	GGAATGAAAATACCACCAAATTATGGCAGGAGATATAGTCAGGCATTTGAAAATAATTATAAGGTAGTGAG
SS544	CTCACTACCTTATAATTATTTTCAAATGCCTGACTATATCTCCTGCCATAATTTGGTGGTATTTTCATTCC
SS545	GCAAAATGACCAGATCCATCCAAATCGCAAAGCCCAGTCAATCTTGCTAAATAACG
SS546	CGTTATTTAGCAAGATTGACTGGGCTTTGCGATTTGGATGGATCTGGTCATTTTGC

Table C-2: Continued

SS547	GAAAATACCACCAAATTATGGCACT**AGA**TATAGTCAGGCATTTG
SS548	CAAATGCCTGACTATA**TCT**AGTGCCATAATTTGGTGGTATTTTC
SS549	GAAAATACCACCAAATTATGGC**AGG**GCCTATAGTCAGGCATTTG
SS550	CAAATGCCTGACTATAGGC**CCT**GCCATAATTTGGTGGTATTTTC
SS583	GAAAATACCACCAAATTATGGCAGG**AGA**TATAGTCAGGCATTTG
SS584	CAAATGCCTGACTATA**TCT**CCTGCCATAATTTGGTGGTATTTTC
SS601	CAAAACCATTCTTATCTTAGGCGAC**GCT**CTGAGTGCGGGTTATGGCATTAACC
SS602	GGTTAATGCCATAACCCGCACTCAG**AGC**GTCGCCTAAGATAAGAATGGTTTTG
SS608	GGCTGGACACAAAAGTCTAATGCAA**GAT**GACCAGATCCATCCAAATGC
SS609	GCATTTGGATGGATCTGGTC**ATC**TTGCATTAGACTTTTGTGTCCAGCC
SS617	CGACAGTCTGAGTGCGGGTTAT**GAA**ATTAACCCCGAACAGGGCTGG
SS618	CCAGCCCTGTTCGGGGTTAAT**TTC**ATAACCCGCACTCAGACTGTCG
SS619	GCAAAATGACCAGATCCATCCAAAT**GAA**AAAGCCCAGTCAATCTTGCTAAATAACG
SS620	CGTTATTTAGCAAGATTGACTGGGCTTT**TTC**ATTTGGATGGATCTGGTCATTTTGC

Table C-3: Chapter 4 primers

Name	**Sequence (SDM primers have codon change in bold)**
SS551	CTACTTACTACTTATCGACCT**SAW**GTGGTGGTCATTGAGCTTGGTGG
SS552	CCACCAAGCTCAATGACCACCAC**WTS**AGGTCGATAAGTAGTAAGTAG

Table C-3: Continued

SS553	CTACTTACTACTTATCGACCT**AGT**GTGGTGGTCATTGAGCTTGGTGG
SS554	CCACCAAGCTCAATGACCACCAC**ACT**AGGTCGATAAGTAGTAAGTAG
SS555	GGACAACCGCCTCAAATGATTCAAAGTAATCTG**KCG**AAATTAATCCAGCACAGCCAAAAG
SS556	CCTTTTGGCTGTGCTGGATTAATTT**CGM**CAGATTACTTTGAATCATTTGAGGCGGTTGTC
SS557	GGACAACCGCCTCAAATGATTCAAAGTAATCTG**GDT**AAATTAATCCAGCACAGCCAAAAG
SS558	CTTTTGGCTGTGCTGGATTAATTT**AHC**CAGATTACTTTGAATCATTTGAGGCGGTTGTCC
SS559	GGAATGAAAATACCACCAAATTATGGC**KCG**GCCTATAGTCAGGCATTTGAAAA
SS560	TTTTCAAATGCCTGACTATAGGC**CGM**GCCATAATTTGGTGGTATTTTCATTCC
SS561	GGAATGAAAATACCACCAAATTATGGC**CAA**GCCTATAGTCAGGCATTTGAAAA
SS562	TTTTCAAATGCCTGACTATAGGC**TTG**GCCATAATTTGGTGGTATTTTCATTCC
SS563	CCCCGAACAGGGCTGGGTC**CAK**TTATTACAAAAACGTCTGGATCAAC
SS564	GTTGATCCAGACGTTTTTGTAATAA**MTG**GACCCAGCCCTGTTCGGGG
SS565	CCCCGAACAGGGCTGGGTC**AMC**TTATTACAAAAACGTCTGGATCAAC
SS566	GTTGATCCAGACGTTTTTGTAATAA**GKT**GACCCAGCCCTGTTCGGGG
SS567	CCTCAAATGATTCAAAGTAATCTGGAA**ARC**TTAATCCAGCACAGCCAAAAGGC
SS568	GCCTTTTGGCTGTGCTGGATTAA**GYT**TTCCAGATTACTTTGAATCATTTGAGG
SS569	CCTCAAATGATTCAAAGTAATCTGGAA**GCA**TTAATCCAGCACAGCCAAAAGGC

Table C-3: Continued

SS570	GCCTTTTGGCTGTGCTGGATTAA**TGC**TTCCAGATTACTTTGAATC ATTTGAGG
SS571	CCTCAAATGATTCAAAGTAATCTGGAA**CAA**TTAATCCAGCACAG CCAAAAGGC
SS572	GCCTTTTGGCTGTGCTGGATTAA**TTG**TTCCAGATTACTTTGAATC ATTTGAGG
SS573	GATCCATCCAAATGCCAAAGCCCAGTCA**VAA**TTGCTAAATAACG CATACCC
SS574	GGGTATGCGTTATTTAGCAA**TTB**TGACTGGGCTTTGGCATTTGG ATGGATC
SS575	CTTGATGGTGTGGCTGGACACAAA**SAT**CTAATGCAAAATGACCA GATCCATCC
SS576	GGATGGATCTGGTCATTTTGCATTAG**ATS**TTTGTGTCCAGCCACA CCATCAAG
SS577	CTTGATGGTGTGGCTGGACACAAA**ACT**CTAATGCAAAATGACC AGATCCATCC
SS578	GGATGGATCTGGTCATTTTGCATTAG**AGT**TTTGTGTCCAGCCAC ACCATCAAG
SS579	CTTGATGGTGTGGCTGGACACAAA**CAA**CTAATGCAAAATGACC AGATCCATCC
SS580	GGATGGATCTGGTCATTTTGCATTAG**TTG**TTTGTGTCCAGCCAC ACCATCAAG
SS581	GGACAACCGCCTCAAATGATTCAA**MAA**AATCTGGAAAAATTAA TCCAGCACAGCC
SS582	GGCTGTGCTGGATTAATTTTTCCAGATT**TTK**TTGAATCATTTGAG GCGGTTGTCC
SS585	CCAGCACAGCCAAAAGGCA**CAK**TCTAAAGTCGTGGTGTTTGGA ATGAAAATACC
SS586	GGTATTTTCATTCCAAACACCACGACTTTAGA**MTG**TGCCTTTTG GCTGTGCTGG
SS587	CCAGCACAGCCAAAAGGCA**ACC**TCTAAAGTCGTGGTGTTTGGA ATGAAAATACC
SS588	GGTATTTTCATTCCAAACACCACGACTTTAGA**GGT**TGCCTTTTG GCTGTGCTGG

Table C-3: Continued

SS589	GGCATTTGAAAATAATTATAAG**AYC**GTGAGTCAAACATATCAG GTTAAGTTG
SS590	CAACTTAACCTGATATGTTTGACTCAC**GRT**CTTATAATTATTTTC AAATGCC
SS591	GGTTAAGTTGTTGCCATTTTTTCTT**SAA**GGTGTGGCTGGACACAA AAGTCTAATGC
SS592	GCATTAGACTTTTGTGTCCAGCCACACC**TTS**AAGAAAAAATGGC AACAACTTAACC
SS666	GGACAACCGCCTCAAATGATTCAA**AAA**AATCTGGAAAAATTAA TCCAGCACAGCC
SS667	GGCTGTGCTGGATTAATTTTTCCAGATT**TTT**TTGAATCATTTGAG GCGGTTGTCC
SS683	CCCCGAACAGGGCTGGGTC**CTG**TTATTACAAAAACGTCTGGATC AAC
SS684	GTTGATCCAGACGTTTTTGTAATAA**CAG**GACCCAGCCCTGTTCG GGG
SS685	CCTCAAATGATTCAAAGTAATCTGGAA**CGC**TTAATCCAGCACAG CCAAAAGGC
SS686	GCCTTTTGGCTGTGCTGGATTAA**GCG**TTCCAGATTACTTTGAAT CATTTGAGG
SS687	GATCCATCCAAATGCCAAAGCCCAGTCA**CTG**TTGCTAAATAACG CATACCC
SS688	GGGTATGCGTTATTTAGCAA**CAG**TGACTGGGCTTTGGCATTTGG ATGGATC
SS689	GGACAACCGCCTCAAATGATTCAA**GCG**AATCTGGAAAAATTAA TCCAGCACAGCC
SS690	GGCTGTGCTGGATTAATTTTTCCAGATT**CGC**TTGAATCATTTGA GGCGGTTGTCC
SS691	CCAGCACAGCCAAAAGGCA**AAT**TCTAAAGTCGTGGTGTTTGGA ATGAAAATACC
SS692	GGTATTTTCATTCCAAACACCACGACTTTAGA**ATT**TGCCTTTTGG CTGTGCTGG
SS693	GGCATTTGAAAATAATTATAAG**CTG**GTGAGTCAAACATATCAG GTTAAGTTG

Table C-3: Continued

SS694	CAACTTAACCTGATATGTTTGACTCAC**CAG**CTTATAATTATTTTC AAATGCC
SS695	GGTTAAGTTGTTGCCATTTTTTCTT**AAT**GGTGTGGCTGGACACAA AAGTCTAATGC
SS696	GCATTAGACTTTTGTGTCCAGCCACACC**ATT**AAGAAAAAATGGC AACAACTTAACC
SS697	CAACCGCCTCAAATGATTCAA**AAA**AATCTG**ACG**AAATTAATCC AGCACAGC
SS698	GCTGTGCTGGATTAATTT**CGT**CAGATT**TTT**TTGAATCATTTGAG GCGGTTG

APPENDIX D. CONSTRUCTION OF SITE DIRECTED MUTANTS

Table D-1: Chapter 3 site directed mutagenesis templates and primers

Plasmid name	Mutation	Template	Mutagenesis primers
pSS196	G17R	PPY1236 (pSS192)	SS541/542
pSS197	T120R	PPY1236 (pSS192)	SS549/550
pSS198	A121R	PPY1236 (pSS192)	SS547/548
pSS199	A165R	PPY1236 (pSS192)	SS545/546
pSS200	G17R/A165R	PPY1310 (pSS196)	SS545/546
pSS201	A121R/A165R	PPY1312 (pSS198)	SS545/546
pSS202	T120R/A165R	PPY1320 (pSS199)	SS549/550
pSS203	G17R/A121R	PPY1312 (pSS198)	SS541/542
pSS204	G17R/T120R	PPY1311 (pSS197)	SS541/542
pSS205	T120R/A121R	PPY1311 (pSS197)	SS583/584
pSS206	S11A	PPY1236 (pSS192)	SS601/602
pSS208	G17R/A165R/T120R	PPY1321 (pSS200)	SS549/550
pSS209	G17R/A165R/A121R	PPY1321 (pSS200)	SS547/548
pSS210	G17R/A165R/N158D	PPY1321 (pSS200)	SS608/609
pSS251	G17E	PPY1310 (pSS196)	SS617/618
pSS252	G17E/A165E	PPY1499 (pSS251)	SS619/620
pSS253	G17E/A165R	PPY1321 (pSS200)	SS617/618
pSS254	G17R/A165E	PPY1310 (pSS196)	SS619/620

Table D-2: Chapter 4 site directed mutagenesis templates and primers

Plasmid names	Mutation	Template	Mutagenesis primers
pSS218-221	N69D/E/H/Q	pSS192	SS551/552
pSS222	N69S	pSS192	SS553/554

Table D-2: continued

pSS225-226	E94A/S	pSS192	SS555/556
pSS227-229	E94D/G/V	pSS192	SS557/558
pSS237-238	T120A/S	pSS192	SS559/560
pSS239	T120Q	pSS192	SS561/562
pSS214-215	A26H/Q	pSS192	SS563/564
pSS216-17	A26N/T	pSS192	SS565/566
pSS230-231	K95N/S	pSS192	SS567/568
pSS232	K95A	pSS192	SS569/570
pSS233	K95Q	pSS192	SS571/572
pSS248-250	I170K/Q/E	pSS192	SS573/574
pSS244-245	S154H/D	pSS192	SS575/576
pSS246	S154T	pSS192	SS577/578
pSS247	S154Q	pSS192	SS579/580
pSS223-224	S91K/Q	pSS192	SS581/582
pSS234-235	K106Q/H	pSS192	SS585/586
pSS236	K106T	pSS192	SS587/588
pSS240-241	V132T/I	pSS192	SS589/590
pSS242-243	D147E/Q	pSS192	SS591/592
pSS258	A26L	pSS192	SS683/684
pSS259	S91A	pSS192	SS685/686
pSS260	K95R	pSS192	SS687/688
pSS261	K106N	pSS192	SS689/690
pSS262	V132L	pSS192	SS691/692
pSS263	D147N	pSS192	SS693/694
pSS264	I170L	pSS192	SS695/696
pSS267	S91K/E94S	pSS223	SS697/698
pSS268	S91K/T120A	pSS237	SS666/667
pSS269	S91K/S154D	pSS245	SS666/667
pSS270	E94S/T120A	pSS226	SS559/560
pSS271	E94S/S154D	pSS226	SS575/576

Table D-2: continued

pSS272	T120A/S154D	pSS245	SS559/560
pSS273	G17R/S91K/A165R	pSS200	SS666/667
pSS274	G17R/E94S/A165R	pSS200	SS555/556
pSS275	G17R/T120A/A165R	pSS200	SS559/560
pSS276	G17R/S154D/A165R	pSS200	SS575/576
pSS277	G17R/T120A/S154D/A165R	pSS275	SS575/576

APPENDIX E. THEORETICAL YIELD CALCULATIONS

Contributed by Nick S. Kruyer (Chapter 2)

E.1 Calculations of stoichiometric theoretical yields of medium-chain chemicals

Yield of medium chain fatty acid (MCFA) based products from glucose. Assume that all carbon from glucose goes into product formation[1]. Gives upper limit of yield.

- **Octanoic Acid**: 8 $C_6H_{12}O_6 \rightarrow 6\ C_8H_{16}O_2$
  - $\frac{6\ moles\ octanoic\ acid}{8\ moles\ glucose} = 75\%\ yield$
- **Decanoic Acid**: 10 $C_6H_{12}O_6 \rightarrow 6\ C_{10}H_{20}O_2$
  - $\frac{6\ moles\ decanoic\ acid}{10\ moles\ glucose} = 60\%\ yield$
- **Dodecanoic Acid:** 12 $C_6H_{12}O_6 \rightarrow 6\ C_{12}H_{24}O_2$
  - $\frac{6\ moles\ dodceanoic\ acd}{12\ moles\ glucose} = 50\%\ yield$
- **Nonane**: 10 $C_6H_{12}O_6 \rightarrow 6\ C_{10}H_{20}O_2 \rightarrow 6\ C_9H_{20}$
  - $\frac{6\ moles\ nonane}{10\ moles\ glucose} = 60\%\ yield$
- **Undecane**: 12 $C_6H_{12}O_6 \rightarrow 6\ C_{12}H_{24}O_2 \rightarrow 6\ C_{11}H_{24}$
  - $\frac{6\ moles\ undecane}{12\ moles\ glucose} = 50\%\ yield$
- **1-Undecene**: 12 $C_6H_{12}O_6 \rightarrow 6\ C_{12}H_{24}O_2 \rightarrow 6\ C_{11}H_{22}$
  - $\frac{6\ moles\ 1-Undecene}{12\ moles\ glucose} = 50\ \%\ yield$
- **Ethyl Octanoic Acid**: 8 $C_6H_{12}O_6 \rightarrow 6\ C_8H_{16}O_2 \rightarrow 6\ C_{10}H_{20}O_2$
  - $\frac{6\ moles\ ethyl\ octanoic\ acid}{8\ moles\ glucose} = 75\%\ yield$
- **Ethyl Decanoic Acid**: 10 $C_6H_{12}O_6 \rightarrow 6\ C_{10}H_{20}O_2 \rightarrow 6\ C_{12}H_{24}O_2$
  - $\frac{6\ moles\ ethyl\ decanoic\ acid}{10\ moles\ glucose} = 60\%\ yield$
- **Ethyl Dodecanoic Acid**: 12 $C_6H_{12}O_6 \rightarrow 6\ C_{12}H_{24}O_2 \rightarrow 6\ C_{14}H_{28}O_2$
  - $\frac{6\ moles\ ethyl\ dodecanoic\ acid}{12\ moles\ glucose} = 50\%\ yield$

- **8-Hydroxy-Octanoic Acid**: 8 $C_6H_{12}O_6 \rightarrow 6\ C_8H_{16}O_3$
  - $\frac{6\ moles\ 8-hydroxy-octanoic\ acid}{8\ moles\ glucose} = 75\%\ yield$
- **10-Hydroxy-Decanoic Acid**: 10 $C_6H_{12}O_6 \rightarrow 6\ C_{10}H_{20}O_3$
  - $\frac{6\ moles\ 10-hydroxy-decanoic\ acid}{10\ moles\ glucose} = 60\%\ yield$
- **12-Hydroxy-Dodecanoic Acid:** 12 $C_6H_{12}O_6 \rightarrow 6\ C_{12}H_{24}O_3$
  - $\frac{6\ moles\ 12-hydroxy-dodecanoic\ acid}{12\ moles\ glucose} = 50\%\ yield$
- **1,8-Octanedioic Acid**: 8 $C_6H_{12}O_6 \rightarrow 6\ C_8H_{14}O_4$
  - $\frac{6\ moles\ 1,8-octanedioic\ acid}{8\ moles\ glucose} = 75\%\ yield$
- **1,9-Nonanedioic Acid**: 9 $C_6H_{12}O_6 \rightarrow 6\ C_9H_{16}O_4$
  - $\frac{6\ moles\ 1,9-nonanedioic\ acid}{9\ moles\ glucose} = 67\%\ yield$
- **1,10-Decanedioic Acid**: 10 $C_6H_{12}O_6 \rightarrow 6\ C_{10}H_{18}O_4$
  - $\frac{6\ moles\ 1,10-decanedioic\ acid}{10\ moles\ glucose} = 60\%\ yield$
- **1,11-Undecanedioic Acid**: 11 $C_6H_{12}O_6 \rightarrow 6\ C_{11}H_{20}O_4$
  - $\frac{6\ moles\ 1,11-undecanedioic\ acid}{11\ moles\ glucose} = 55\%\ yield$
- **1,12-Dodecanedioic Acid**: 12 $C_6H_{12}O_6 \rightarrow 6\ C_{12}H_{22}O_4$
  - $\frac{6\ moles\ 1,12-dodecanedioic\ acid}{12\ moles\ glucose} = 50\%\ yield$
- **Octanol**: 8 $C_6H_{12}O_6 \rightarrow 6\ C_8H_{18}O$
  - $\frac{6\ moles\ octanol}{8\ moles\ glucose} = 75\%\ yield$
- **Decanol**: 10 $C_6H_{12}O_6 \rightarrow 6\ C_{10}H_{22}O$
  - $\frac{6\ moles\ decanol}{10\ moles\ glucose} = 60\%\ yield$
- **Undecanol**: 12 $C_6H_{12}O_6 \rightarrow 6\ C_{12}H_{24}O_2 \rightarrow 6\ C_{11}H_{24}O$
  - $\frac{6\ moles\ undecanol}{12\ moles\ glucose} = 50\%\ yield$
- **Dodecanol**: 12 $C_6H_{12}O_6 \rightarrow 6\ C_{12}H_{26}O$
  - $\frac{6\ moles\ dodecanol}{12\ moles\ glucose} = 50\%\ yield$
- **2-Undecanone**: 12 $C_6H_{12}O_6 \rightarrow 6\ C_{12}H_{24}O_2 \rightarrow 6\ C_{11}H_{22}O$

- $\frac{6\ moles\ 2-undecanone}{12\ moles\ glucose} = 50\%\ yield$

- **2-Tridecanone**: 14 $C_6H_{12}O_6 \rightarrow 6\ C_{14}H_{28}O_2 \rightarrow 6\ C_{13}H_{26}O$
  - $\frac{6\ moles\ 2-tridecanone}{14\ moles\ glucose} = 43\%\ yield$

E.2 Calculations of theoretical yield of medium-chain chemicals

Maximum theoretical yield calculated using the COBRA 3.0 toolbox implemented in Matlab[2]. Pathways were modeled in *Escherichia coli* using genomic model iAF1260b[3] and *Yarrowia lipolytica* model iNL895[4]. Glucose uptake rate (EX_glc_D_e) was fixed at -10, and O_2 uptake rate (EX_o2_e) was fixed at -15 for all pathway simulations.

To add reversed β-oxidation (rBOX) to the model, directionality of the homologous reactions was changed from fatty acid degradation to chain elongation. All involved reactions were restricted to be unidirectional. In order to promote production of the desired chain length, reactions that promote chain elongation above desired length were removed from the model. Involved reactions:

- **ACAT1r**: Acetyl-CoA C-acetyltransferase
  - Removed reverse reaction
- **ECOAH(1-4)**: 3-hydroxyacyl-CoA dehydratase
  - Removed reverse reaction
- **HACD(1-4)i**: 3-hydroxyacyl-CoA dehydrogenase
  - Added reverse reaction
  - Removed forward reaction
- **ACOAD(1-4)f**: Acyl-CoA dehydrogenase
  - Removed forward reaction
- **KAT(2-4)**: 3-ketoacyl-CoA thiolase
  - Added reverse reaction
  - Removed forward reaction

For simulating production of MCFA derived compounds from the fatty acid biosynthesis (FAS) pathway, reactions associated with the fadD and fadE genes were removed from the model. The involved reactions:

- fadD: **FACOAL(60-181)t2pp**: Fatty-acid-CoA Ligase
- fadE: **ACOAD(1-8)f**: Acyl-CoA Dehydrogenase

The differences in rBOX vs FAS production comes from production of the product precursors (bolded in reaction schemes). Downstream production of final products remains the same. The following equations were added for production of MCFA derived products for the *E. coli* model:

Octanoic Acid

OCTAe: **octa[c]** → octa[e]

EX_octa_e: octa[e] → Exchange Reaction

Decanoic Acid

DCAe: **dca[c]** → dca[e]

EX_dca_e: dca[e] → Exchange Reaction

Dodecanoic Acid

DDCAe: **ddca[c]** → ddca[e]

EX_ddca_e: ddca[e] → Exchange Reaction

Nonane[5]

acylRED: **dcacoa[c]** + nadph[c] + h[c] → dcaALD[c] + coa[c] + nadp[c]
Acyl-CoA Reductase

ALDoxy: dcaALD[c] + o2[c] + nadh[c] + h[c] → non[c] + for[c] + h2o[c] + nad[c]
Aldehyde Deoxygenase

NONe: non[c] → non[e]

EX_non[e]: non[e] →
Exchange Reaction

Undecane[5]

acylRED: **ddcacoa[c]** + nadph[c] + h[c] → ddcaALD[c] + coa[c] + nadp[c]
Acyl-CoA Reductase

ALDoxy: ddcaALD[c] + o2[c] + nadh[c] + h[c] → und[c] + for[c] + h2o[c] + nad[c]
Aldehyde Deoxygenase

UNDe: und[c] → und[e]

EX_und[e]: und[e] →
Exchange Reaction

1-Undecene[6]

FADC: **ddca[c]** + o2[c] → 1und[c] + co2[c] + h2o2[c]
Fatty Acid Decarboxylase

1UNDe: 1und[c] → 1und[e]

EX_1und[e]: 1und[e] →
Exchange Reaction

Ethyl Octanoic Acid[7]

ACEAO: **occoa[c]** + nad[c] + etoh[c] → octaEE[c] + coa[c] + nadh[c]
Acyl-CoA: Ethanol O-acyltransferase

octaEEe: octaEE[c] → octaEE[e]

EX_octaEE[e]: octaEE[e] →
Exchange Reaction

Ethyl Decanoic Acid[5]

ACEAO: **dcacoa[c]** + nad[c] + etoh[c] → dcaEE[c] + coa[c] + nadh[c]
Acyl-CoA: Ethanol O-acyltransferase

dcaEEe: dcaEE[c] → dcaEE[e]

EX_dcaEE[e]: dcaEE[e] →
Exchange Reaction

Ethyl Dodecanoic Acid[5]

ACEAO: **ddcacoa[c]** + nad[c] + etoh[c] → ddcaEE[c] + coa[c] + nadh[c]
Acyl-CoA: Ethanol O-acyltransferase

ddcaEEe: ddcaEE[c] → ddcaEE[e]

EX_ddcaEE[e]: ddcaEE[e] →
Exchange Reaction

8-Hydroxy Octanoic Acid[8]

wOHOCTA: **octa[c]** + o2[c] + 2 h[c] + nadh[c] → octaOH[c] + h2o[c] + nad[c]
Carboxylic Acid Mono-oxygenase

octaOHe: octaOH[c] → octaOH[e]

EX_octaOH[e]: octaOH[e] →
Exchange Reaction

10-Hydroxy Decanoic Acid[8]

wOHDCA: **dca[c]** + o2[c] + h[c] + nadh[c] → dcaOH[c] + h2o[c] + nad[c]
Carboxylic Acid Mono-oxygenase

dcaOHe: dcaOH[c] → dcaOH[e]

EX_dcaOH[e]: dcaOH[e] →
Exchange Reaction

12-Hydroxy Dodecanoic Acid[8]

wOHDCA: **ddca[c]** + o2[c] + h[c] + nadh[c] → ddcaOH[c] + h2o[c] + nad[c]
Carboxylic Acid Mono-oxygenase

ddcaOHe: ddcaOH[c] → ddcaOH[e]

EX_ddcaOH[e]: ddcaOH[e] →
Exchange Reaction

1,8-Octanedioic Acid[8]

wOHOCTA: **octa[c]** + o2[c] + h[c] + nadh[c] → octaOH[c] + h2o[c] + nad[c]
Carboxylic Acid Mono-oxygnease

octaAHD: octaOH[c] + nad[c] → octaALD[c] + nadh[c] + h[c]
Alcohol Dehydrogenase

octaALDH: octaALD[c] + h2o[c] + nad[c] → octaDiCOOH[c] + nadh[c] + h[c]
Aldehyde Dehydrogenase

octaDiCOOHe: octaDiCOOH[c] → octaDiCOOH[e]

EX_octaDiCOOH[e]: octaDiCOOH[e] →
Exchange Reaction

1,9-Nonanedioic Acid[9]

BioC: **malACP[c]** + amet[c] → malACPME[c] + ahcys[c]

FAS5: malACPME[c] + acACP[c] + h[c] → c5ACPME[c] + ACP[c] + co2[c]

FAS7: c5ACPME[c] + acACP[c] + h[c] → c7ACPME[c] + ACP[c] + co2[c]

FAS9: c7ACPME[c] + acACP[c] + h[c] → c9ACPME[c] + ACP[c] + co2[c]

TesA: c9ACPME[c] + h2o[c] → C9ME[c] + ACP[c] + h[c]

HYD: C9ME[c] + h2o[c] → C9Di[c]

C9DIe: C9Di[c] → C9Di[e]

EX_C9Di[e]: C9Di[e] → Exchange Reaction

1,10-Decanedioic Acid[8]

wOHDCA: **dca[c]** + o2[c] + h[c] + nadh[c] → dcaOH[c] + h2o[c] + nad[c] Carboxylic Acid Mono-oxygenase

dcaAHD: dcaOH[c] + nad[c] → dcaALD[c] + nadh[c] + h[c] Alcohol Dehydrogenase

dcaALDH: dcaALD[c] + h2o[c] + nad[c] → dcaDiCOOH[c] + nadh[c] + h[c] Aldehyde Dehydrogenase

dcaDiCOOHe: dcaDiCOOH[c] → dcaDiCOOH[e]

EX_dcaDiCOOH[e]: dcaDiCOOH[e] → Exchange Reaction

1,11-Undecanedioic Acid[9]

BioC: **malACP[c]** + amet[c] → malACPME[c] + ahcys[c]

FAS5: malACPME[c] + acACP[c] + h[c] → c5ACPME[c] + ACP[c] + co2[c]

FAS7: c5ACPME[c] + acACP[c] + h[c] → c7ACPME[c] + ACP[c] + co2[c]

FAS9: c7ACPME[c] + acACP[c] + h[c] → c9ACPME[c] + ACP[c] + co2[c]

FAS11: c9ACPME[c] + acACP[c] + h[c] → c11ACPME[c] + ACP[c] + co2[c]

TesA: c11ACPME[c] + h2o[c] → C11ME[c] + ACP[c] + h[c]

HYD: C11ME[c] + h2o[c] → C11Di[c]

C11DIe: C11Di[c] → C11Di[e]

EX_C11Di[e]: C11Di[e] →
Exchange Reaction

1,12-Dodecanedioic Acid[8]

wOHDCA: **ddca[c]** + o2[c] + h[c] + nadh[c] → ddcaOH[c] + h2o[c] + nad[c]
Carboxylic Acid Mono-oxygenase

dcaAHD: ddcaOH[c] + nad[c] → ddcaALD[c] + nadh[c] + h[c]
Alcohol Dehydrogenase

dcaALDH: ddcaALD[c] + h2o[c] + nad[c] → ddcaDiCOOH[c] + nadh[c] + h[c]
Aldehyde Dehydrogenase

dcaDiCOOHe: ddcaDiCOOH[c] → ddcaDiCOOH[e]

EX_dcaDiCOOH[e]: ddcaDiCOOH[e] →
Exchange Reaction

Octanol[10, 11]

CAR: **octa[c]** + nadph[c] + atp[c] + 2 h[c] → octaALD[c] + nadp[c] + amp[c] + h2o[c] + ppi[c]
Carboxylic Acid Reductase

octaADH: octaALD[c] + nadph[c] → octaOL[c] + nadp[c]
Aldehyde Reductase

octaOLe: octaOL[c] → octaOL[e]

EX_octaOL[e]: octaOL[e] →
Exchange Reaction

Decanol[10, 11]

CAR: **dca[c]** + nadph[c] + atp[c] + 2 h[c] → dcaALD[c] + nadp[c] + amp[c] + h2o[c] + ppi[c]
Carboxylic Acid Reductase

dcaADH: dcaALD[c] + nadph[c] → dcaOL[c] + nadp[c]
Aldehyde Reductase

dcaOLe: dcaOL[c] → dcaOL[e]

EX_dcaOL[e]: dcaOL[e] →
Exchange Reaction

Undecanol[12]

αDOX: **ddca[c]** + o2[c] → udcaALD[c] + co2[c] + h2o[c]
α-Dioxygenase

udcaADH: udcaALD[c] + nadph[c] → udcaOL[c] + nadp[c]
Aldehyde Reductase

dcaOLe: udcaOL[c] → udcaOL[e]

EX_dcaOL[e]: udcaOL[e] →
Exchange Reaction

Dodecanol[10, 11]

CAR: **ddca[c]** + nadph[c] + atp[c] + 2 h[c] → ddcaALD[c] + nadp[c] + amp[c] + h2o[c] + ppi[c]
Carboxylic Acid Reductase

dcaADH: ddcaALD[c] + nadph[c] → ddcaOL[c] + nadp[c]
Aldehyde Reductase

dcaOLe: ddcaOL[c] → ddcaOL[e]

EX_dcaOL[e]: ddcaOL[e] →
Exchange Reaction

2-Undecanone[13]

kaTES: **3oddcoa[c]** + h2o[c] → BketoFA[c] + h[c] + coa[c]
Keto-acyl Thioesterase

ketoDC: BketoFA[c] → 2unnone[c] + co2[c]
Keto-acyl Decarboxylase

2UNNONEe: 2unnone[c] → 2unnone[e]

EX_2unnone[e]: 2unnone[e] →
Exchange Reaction

2-Tridecanone[13]

kaTES: **3otdcoa[c]** + h2o[c] → BketoFA[c] + h[c] + coa[c]
Keto-acyl Thioesterase

ketoDC: BketoFA[c] → 2trinone[c] + co2[c]
Keto-acyl Decarboxylase

2UNNONEe: 2trinone[c] → 2trinone[e]

EX_2trinone[e]: 2trinone[e] →
Exchange Reaction

The highest titers found in literature for decanoic acid and decanol were both achieved using the fatty acid biosynthesis pathway in *Y. lipolytica*. Therefore, production of these two compounds was modeled in *Y. lipolytica* using the following reactions:

Decanoic Acid[14]

DCAe: **s_0575[c_02]** → s_0577[c_05]

Decanol[14]

acyRED: **s_0582[c_02]** + s_1087[c_02] → dcaOH[c_02] + s_1082[c_02] + s_0514[c_02]
Dual Acyl-CoA Reductase

dcaOHe: dcaOH[c_02] → dcaOH[c_05]

EX_dcaOH[c_05]: dcaOH[c_05] →
Exchange Reaction

E.3 References

1 Zhang, H., Li, Z., Pereira, B. & Stephanopoulos, G. Engineering E. coli-E. coli cocultures for production of muconic acid from glycerol. *Microb. Cell Fact.* **14**, 134 (2015).

2 Schellenberger, J. et al. Quantitative prediction of cellular metabolism with constraint-based models: the COBRA Toolbox v2.0. *Nat. Protoc*. **6**, 1290-1307 (2011).

3 Feist, A.M. et al. Model-driven evaluation of the production potential for growth-coupled products of Escherichia coli. *Metab. Eng*. **12**, 173-186 (2010).

4 Loira, N., Dulermo, T., Nicaud, J.M. & Sherman, D.J. A genome-scale metabolic model of the lipid-accumulating yeast Yarrowia lipolytica. *BMC Syst. Biol.* **6**, 35 (2012).

5 Choi, Y.J. & Lee, S.Y. Microbial production of short-chain alkanes. *Nature* **502**, 571-574 (2013).

6 Rui, Z., Harris, N.C., Zhu, X.J., Huang, W. & Zhang, W.J. Discovery of a Family of Desaturase-Like Enzymes for 1-Alkene Biosynthesis. *ACS Catal*. **5**, 7091-7094 (2015).

7 Lian, J. & Zhao, H. Reversal of the beta-oxidation cycle in Saccharomyces cerevisiae for production of fuels and chemicals. *ACS Synth. Biol*. **4**, 332-41 (2015).

8 Clomburg, J.M. et al. Integrated engineering of beta-oxidation reversal and omega-oxidation pathways for the synthesis of medium chain omega-functionalized carboxylic acids. *Metab. Eng*. **28**, 202-212 (2015).

9 Haushalter, R.W. et al. Production of Odd-Carbon Dicarboxylic Acids in Escherichia coli Using an Engineered Biotin–Fatty Acid Biosynthetic Pathway. *J. Am. Chem. Soc*. **139**, 4615-4618 (2017)

10 Akhtar, M.K., Turner, N.J. & Jones, P.R. Carboxylic acid reductase is a versatile enzyme for the conversion of fatty acids into fuels and chemical commodities. *Proc Nat Acad Sci USA* **110**, 87-92 (2013).

11 Dellomonaco, C., Clomburg, J.M., Miller, E.N. & Gonzalez, R. Engineered reversal of the beta-oxidation cycle for the synthesis of fuels and chemicals. *Nature* **476**, 355-359 (2011).

12 Cao, Y.X. et al. Biosynthesis of odd-chain fatty alcohols in Escherichia coli. *Metab. Eng.* **29**, 113-123 (2015).

13 Goh, E.B., Baidoo, E.E.K., Keasling, J.D. & Beller, H.R. Engineering of Bacterial Methyl Ketone Synthesis for Biofuels. *Appl. Environ. Microbiol*. **78**, 70-80 (2012).

14 Rutter, C.D. & Rao, C.V. Production of 1-decanol by metabolically engineered Yarrowia lipolytica. *Metab. Eng*. **38**, 139-147 (2016).

APPENDIX F. DNA SEQUENCES

F.1 Chapter 1 sequences

F.1.1 Abies grandis geraniol diphosphate synthase (E. coli codon-optimized, N-term signal peptide removed):

ATGGAATTTGACTTCAACAAATACATGGACTCCAAAGCGATGACGGTAAATG
AAGCACTGAACAAAGCGATCCCTCTGCGTTATCCGCAGAAAATCTACGAAAG
CATGCGTTACAGCCTGCTGGCAGGCGGCAAGCGTGTTCGTCCGGTTCTGTGTA
TTGCCGCATGTGAACTGGTAGGTGGTACCGAAGAACTGGCGATCCCGACCGC
GTGCGCAATTGAAATGATCCACACGATGTCCCTGATGCACGATGATCTGCCG
TGTATCGACAACGACGATCTGCGTCGCGGTAAACCGACTAACCACAAAATTT
TCGGTGAGGATACCGCAGTGACTGCTGGTAACGCACTGCACTCTTACGCCTTC
GAGCATATCGCGGTTTCTACTTCTAAAACCGTTGGTGCTGACCGCATCCTGCG
TATGGTGTCCGAGCTGGGTCGTGCTACTGGCTCTGAAGGTGTTATGGGTGGTC
AGATGGTAGACATCGCATCCGAAGGCGATCCGTCTATCGACCTGCAGACCCT
GGAATGGATTCACATCCACAAAACCGCAATGCTGCTGGAATGCTCCGTTGTTT
GCGGTGCAATCATTGGCGGTGCCAGCGAAATCGTAATCGAACGTGCCCGTCG
CTACGCCCGCTGTGTTGGTCTGCTGTTCCAGGTAGTTGATGACATTCTGGACG
TAACTAAAAGCAGCGACGAACTGGGTAAGACTGCGGGCAAGGACCTGATCTC
TGATAAAGCCACCTACCCAAAGCTGATGGGTCTGGAAAAGGCCAAGGAGTTC
TCCGATGAACTGCTGAACCGTGCGAAGGGTGAACTGTCCTGCTTCGACCCAG
TTAAAGCCGCTCCGCTGCTGGGCCTGGCAGACTACGTGGCATTTCGTCAGAAT
TAA

F.1.2 Picea abies geraniol diphosphate synthase (E. coli codon-optimized, N-term signal peptide removed):

ATGGAGTTCGACTTCGACAAATACATGCACTCTAAAGCGATCGCGGTTAACG
AAGCGCTGGACAAAGTTATCCCGCCGCGTTACCCGCAGAAAATCTACGAATC
TATGCGTTACTCTCTGCTGGCGGGTGGTAAACGTGTTCGTCCGATCCTGTGCA
TCGCGGCGTGCGAACTGATGGGTGGTACCGAAGAACTGGCGATGCCGACCGC
GTGCGCGATCGAAATGATCCACACCATGTCTCTGATCCACGACGACCTGCCG
TACATCGACAACGACGACCTGCGTCGTGGTAAACCGACCAACCACAAAGTTT
TCGGTGAAGACACCGCGATCATCGCGGGTGACGCGCTGCTGTCTCTGGCGTT
CGAACACGTTGCGGTTTCTACCTCTCGTACCCTGGGTACCGACATCATCCTGC
GTCTGCTGTCTGAAATCGGTCGTGCGACCGGTTCTGAAGGTGTTATGGGTGGT
CAGGTTGTTGACATCGAATCTGAAGGTGACCCGTCTATCGACCTGGAAACCC
TGGAATGGGTTCACATCCACAAAACCGCGGTTCTGCTGGAATGCTCTGTTGTT
TGCGGTGCGATCATGGGTGGTGCGTCTGAAGACGACATCGAACGTGCGCGTC
GTTACGCGCGTTGCGTTGGTCTGCTGTTCCAGGTTGTTGACGACATCCTGGAC
GTTTCTCAGTCTTCTGAAGAACTGGGTAAAACCGCGGGTAAAGACCTGATCT
CTGACAAAGCGACCTACCCGAAACTGATGGGTCTGGAAAAAGCGAAAGAGT
TCGCGGACGAACTGCTGAACCGTGGTAAACAGGAACTGTCTTGCTTCGACCC
GACCAAAGCGGCGCCGCTGTTCGCGCTGGCGGACTACATCGCGTCTCGTCAG
AACTAA

F.1.3 Pinus taeda geraniol diphosphate synthase (E. coli codon-optimized, N-term signal peptide removed)

ATGGGTTTCGACCTGCGTCGTTACTGGAACTCTGTTCGTAAAGAAATGGAAG
CGGAACTGGAAAAAGCGGTTCCGCTGCGTTACCCGGAAAAAATCCACGTTTC
TATGCGTTACTCTGTTCTGGCGGAAAAAGCGAAACGTGGTCCGCCGATGATG
TGCATCGCGGCGTGCGAAGCGGTTGGTGGTTCTAAAGAACAGGCGCTGCCGA
CCGCGTGCGCGCTGGAAATGGTTCACACCGCGTCTCTGATCCACGACGACCT
GCCGTGCATGGACGACGACCCGATCCGTCGTGGTCAGCCGTCTAACCACACC
GTTTTCGGTGTTGACATGGCGGTTCTGGCGGGTGACGCGCTGTTCCCGCTGGG
TTTCGAATACATCGTTTCTTCTACCTCTGACGTTCCGCCGGACCGTATCGTTAA
AGTTATCGCGCAGATCGCGAAAACCGTTGGTTCTCAGGGTATGGTTGCGGGT
CAGTACATGGACCTGGGTTCTACCGGTAACAACAAATTCGACCCGTCTCTGGT
TGGTTTCATCCACGAACGTAAATTCGGTATCATGGCGGAATGCTCTGCGGTTA
CCGGTGGTATCATCGGTGGTGCGACCGACGAAGACGTTGAACGTCTGCGTAA
CTACGGTCGTTCTGTTGGTGTTCTGTACCAGATCGCGGACGACATCCTGGAAG
AAGAAGCGCTGCTGGAAGGTAACGGTGGTGCGATCAAAAAAAACAAAACCT
CTTACCCGCGTGCGTACGGTCTGGAAAAAGCGCGTGAAATCGCGGAAACCCT
GCGTTCTAACGCGAAAATGGAACTGTCTGGTTTCGACCGTGGTAAATCTGCG
CCGCTGTACTCTTTCGTTGACTACGCGATCGACCGTTCTTTCGACGTTTAA

F.1.4 Abies grandis pinene synthase (E. coli codon-optimized, N-term signal peptide removed)

ATGCGTCGTGGTAAATCTATCACCCCGTCTATCTCTATGTCTTCTACCACCGTT
GTTACCGACGACGGTGTTCGTCGTCGTATGGGTGACTTCCACTCTAACCTGTG
GGACGACGACGTTATCCAGTCTCTGCCGACCGCGTACGAAGAAAAATCTTAC
CTGGAACGTGCGGAAAAACTGATCGGTGAAGTTAAAAACATGTTCAACTCTA

TGTCTCTGGAAGACGGTGAACTGATGTCTCCGCTGAACGACCTGATCCAGCG

TCTGTGGATCGTTGACTCTCTGGAACGTCTGGGTATCCACCGTCACTTCAAAG

ACGAAATCAAATCTGCGCTGGACTACGTTTACTCTTACTGGGGTGAAAACGG

TATCGGTTGCGGTCGTGAATCTGTTGTTACCGACCTGAACTCTACCGCGCTGG

GTCTGCGTACCCTGCGTCTGCACGGTTACCCGGTTTCTTCTGACGTTTTCAAA

GCGTTCAAAGGTCAGAACGGTCAGTTCTCTTGCTCTGAAAACATCCAGACCG

ACGAAGAAATCCGTGGTGTTCTGAACCTGTTCCGTGCGTCTCTGATCGCGTTC

CCAGGTGAAAAAATCATGGACGAAGCGGAAATCTTCTCTACCAAATACCTGA

AAGAAGCGCTGCAGAAAATCCCGGTTTCTTCTCTGTCTCGTGAAATCGGTGAC

GTTCTGGAATACGGTTGGCACACCTACCTGCCGCGTCTGGAAGCGCGTAACT

ACATCCAGGTTTTCGGTCAGGACACCGAAAACACCAAATCTTACGTTAAATC

TAAAAAACTGCTGGAACTGGCGAAACTGGAATTTAACATCTTCCAGTCTCTG

CAGAAACGTGAACTGGAATCTCTGGTTCGTTGGTGGAAAGAATCTGGTTTCC

CGGAAATGACCTTCTGCCGTCACCGTCACGTTGAATACTACACCCTGGCGTCT

TGCATCGCGTTCGAACCGCAGCACTCTGGTTTCCGTCTGGGTTTCGCGAAAAC

CTGCCACCTGATCACCGTTCTGGACGACATGTACGACACCTTCGGTACCGTTG

ACGAACTGGAACTGTTCACCGCGACCATGAAACGTTGGGACCCGTCTTCTAT

CGACTGCCTGCCGGAATACATGAAAGGTGTTTACATCGCGGTTTACGACACC

GTTAACGAAATGGCGCGTGAAGCGGAAGAAGCGCAGGGTCGTGACACCCTG

ACCTACGCGCGTGAAGCGTGGGAAGCGTACATCGACTCTTACATGCAGGAAG

CGCGTTGGATCGCGACCGGTTACCTGCCGTCTTTCGACGAATACTACGAAAA

CGGTAAAGTTTCTTGCGGTCACCGTATCTCTGCGCTGCAGCCGATCCTGACTA

TGGACATCCCGTTCCCGGACCACATCCTGAAAGAAGTTGACTTCCCGTCTAA

ACTGAACGACCTGGCGTGCGCGATCCTGCGTCTGCGTGGTGACACCCGTTGC

TACAAAGCGGACCGTGCGCGTGGTGAAGAAGCGTCTTCTATCTCTTGCTACAT

GAAAGACAACCCAGGTGTTTCTGAAGAAGACGCGCTGGACCACATCAACGC

GATGATCTCTGACGTTATCAAAGGTCTGAACTGGGAACTGCTGAAACCGGAC

ATCAACGTTCCGATCTCTGCGAAAAAACACGCGTTCGACATCGCGCGTGCGT

TCCACTACGGTTACAAATACCGTGACGGTTACTCTGTTGCGAACGTTGAAACC

AAATCTCTGGTTACCCGTACCCTGCTGGAATCTGTTCCGCTGTAA

F.1.5 Picea abies pinene synthase (E. coli codon-optimized, N-term signal peptide removed)

ATGCGTCGTATGGGTGACTTCCACTCTAACCTGTGGAACGACGACTTCATCCA

GTCTCTGTCTACCTCTTACGGTGAACCGTCTTACCGTGAACGTGCGGAACGTC

TGATCGGTGAAGTTAAAAAAATGTTCAACTCTATGTCTTCTGAAGACGGTGA

ACTGATCTCTCCGCACAACGACCTGATCCAGCGTGTTTGGATGGTTGACTCTG

TTGAACGTCTGGGTATCGAACGTCACTTCAAAAACGAAATCAAATCTGCGCT

GGACTACGTTTACTCTTACTGGTCTGAAAAAGGTATCGGTTGCGGTCGTGAAT

CTGTTGTTGCGGACCTGAACTCTACCGCGCTGGGTCTGCGTACCCTGCGTCTG

CACGGTTACGCGGTTTCTGCGGACGTTCTGAACCTGTTCAAAGACCAGAACG

GTCAGTTCGCGTGCTCTCCGTCTCAGACCGAAGAAGAAATCCGTTCTGTTCTG

AACCTGTACCGTGCGTCTCTGATCGCGTTTCCGGGTGAAAAAGTTATGGAAG

AAGCGGAAATCTTCTCTGCGAAATACCTGGAAGAAGCGCTGCAGAAAATCTC

TGTTTCTTCTCTGTCTCAGGAAATCCGTGACGTTCTGGAATACGGTTGGCACA

CCTACCTGCCGCGTATGGAAGCGCGTAACCACATCGACGTTTTCGGTCAGGA

CACCCAGAACTCTAAATCTTGCATCAACACCGACAAACTGCTGGAACTGGCG

AAACTGGAGTTCAACATCTTCCACTCTCTGCAGAAACGTGAACTGGAATACC
TGGTTCGTTGGTGGAAAGACTCTGGTTCTCCGCAGATGACCTTCGGTCGTCAC
CGTCACATCGAATACTACACCCTGGCGTCTTGCATCGCGTTCGAACCGCAGC
ACTCTGGTTTCCGTCTGGGTTTCGCGAAAACCTGCCACATCATCACCATCCTG
GACGACATGTACGACACCTTCGGTACCGTTGACGAACTGGAACTGTTCACCG
CGGCGATGAAACGTTGGGACCCGTCTGCGGCGGACTGCCTGCCGGAATACAT
GAAAGTTATGTACATGATCGTTTACGACACCGTTAACGAAATGTGCCAGGAA
GCGGAAAAAGCGCAGGGTCGTGACACCCTGGACTACGCGCGTCAGGCGTGG
GAAGACTACCTGGACTCTTACATGCAGGAAGCGAAATGGATCGCGACCGGTT
ACCTGCCGACCTTCGAAGAATACTACGAAAACGGTAAAGTTTCTTCTGGTCA
CCGTGTTGCGGCGCTGCAGCCGATCCTGACTATGGACATCCCGTTCCCGCCGC
ACATCCTGAAAGAAGTTGACTTCCCGTCTAAACTGTCTGACCTGGCGTGCGCG
ATCCTGCGTCTGCGTGGTGACACCCGTTGCTACAAAGCGGACCGTGCGCGTG
GTGAAGAAGCGTCTTCTATCTCTTGCTACATGAAAGACAATCCGGGTGCGAC
CGAAGAAGACGCGCTGGACCACATCAACGCGATGATCTCTGACGTTATCCGT
GGTCTGAACTGGGAACTGCTGAAACCGAACTCTTCTGTTCCGATCTCTTCTAA
AAAACACGTTTTCGACATCTCTCGTGCGTTCCACTACGGTTACAAATACCGTG
ACGGTTACTCTGTTGCGAACATCGAAACCAAATCTCTGGTTAAACGTACCGTT
ATCGACCCGGTTACCCTGTAA

F.1.6 Pinus taeda pinene synthase (E. coli codon-optimized, N-term signal peptide removed)

ATGGCGCGTCGTACCCTGTTCGGTTTCTCTCACGAACTGAAAGCGATCCACTC
TACCGTTCCGAACCTGGGTATGTGCCGTGGTGGTAAATCTATCGCGCCGTCTA

TGTCTATGTCTTCTACCACCTCTGTTTCTAACGAAGACGGTGTTCCGCGTCGT

ATCGCGGGTCACCACTCTAACCTGTGGGACGACGACTCTATCGCGTCTCTGTC

TACCTCTTACGAAGCGCCGTCTTACCGTAAACGTGCGGACAAACTGATCGGT

GAAGTTAAAAACATCTTCGACCTGATGTCTGTTGAAGACGGTGTTTTCACCTC

TCCGCTGTCTGACCTGCACCACCGTCTGTGGATGGTTGACTCTGTTGAACGTC

TGGGTATCGACCGTCACTTCAAAGACGAAATCAACTCTGCGCTGGACCACGT

TTACTCTTACTGGACCGAAAAAGGTATCGGTCGTGGTCGTGAATCTGGTGTTA

CCGACCTGAACTCTACCGCGCTGGGTCTGCGTACCCTGCGTCTGCACGGTTAC

ACCGTTTCTTCTCACGTTCTGGACCACTTCAAAAACGAAAAAGGTCAGTTCAC

CTGCTCTGCGATCCAGACCGAAGGTGAAATCCGTGACGTTCTGAACCTGTTCC

GTGCGTCTCTGATCGCGTTCCCTGGTGAAAAAATCATGGAAGCGGCGGAAAT

CTTCTCTACCATGTACCTGAAAGACGCGCTGCAGAAAATCCCGCCGTCTGGTC

TGTCTCAGGAAATCGAATACCTGCTGGAATTTGGTTGGCACACCAACCTGCC

GCGTATGGAAACCCGTATGTACATCGACGTTTTCGGTGAAGACACCACCTTC

GAAACCCCGTACCTGATCCGTGAAAAACTGCTGGAACTGGCGAAACTGGAAT

TTAACATCTTCCACTCTCTGGTTAAACGTGAACTGCAGTCTCTGTCTCGTTGGT

GGAAAGACTACGGTTTCCCGGAAATCACCTTCTCTCGTCACCGTCACGTTGAA

TACTACACCCTGGCGGCGTGCATCGCGAACGACCCGAAACACTCTGCGTTCC

GTCTGGGTTTCGGTAAAATCTCTCACATGATCACCATCCTGGACGACATCTAC

GACACCTTCGGTACAATGGAAGAACTGAAACTGCTGACCGCGGCGTTCAAAC

GTTGGGACCCGTCTTCTATCGAATGCCTGCCGGACTACATGAAAGGTGTTTAC

ATGGCGGTTTACGACAACATCAACGAAATGGCGCGTGAAGCGCAGAAAATCC

AGGGTTGGGACACCGTTTCTTACGCGCGTAAATCTTGGGAAGCGTTCATCGGT

GCGTACATCCAGGAAGCGAAATGGATCTCTTCTGGTTACCTGCCGACCTTCGA
CGAATACCTGGAAAACGGTAAAGTTTCTTTCGGTTCTCGTATCACCACCCTGG
AACCGATGCTGACCCTGGGTTTCCCGCTGCCGCCGCGTATCCTGCAGGAAAT
CGACTTCCCGTCTAAATTCAACGACCTGATCTGCGCGATCCTGCGTCTGAAAG
GTGACACCCAGTGCTACAAAGCGGACCGTGCGCGTGGTGAAGAAGCGTCTGC
GGTTTCTTGCTACATGAAAGACCACCCAGGTATCACCGAAGAAGACGCGGTT
AACCAGGTTAACGCGATGGTTGACAACCTGACCAAAGAACTGAACTGGGAAC
TGCTGCGTCCGGACTCTGGTGTTCCGATCTCTTACAAAAAAGTTGCGTTCGAC
ATCTGCCGTGTTTTCCACTACGGTTACAAATACCGTGACGGTTTCTCTGTTGC
GTCTATCGAAATCAAAAACCTGGTTACCCGTACCGTTGTTGAAACCGTTCCGC
TGTAA

F.2 Chapter 3 sequences

F.2.1 Acinetobacter baylyi thioesterase (non-codon optimized, N-terminal signal peptide removed)

ATGGGCAAAACCATTCTTATCTTAGGCGACAGTCTGAGTGCGGGTTATGGCA
TTAACCCCGAACAGGGCTGGGTCGCTTTATTACAAAAACGTCTGGATCAACA
ATTTCCCAAGCAGCATAAAGTCATTAATGCCAGTGTAAGTGGGGAAACCACC
AGTGGTGCTTTAGCTCGTTTACCCAAACTACTTACTACTTATCGACCTAATGT
GGTGGTCATTGAGCTTGGTGGTAATGATGCATTAAGAGGACAACCGCCTCAA
ATGATTCAAAGTAATCTGGAAAAATTAATCCAGCACAGCCAAAAGGCAAAAT
CTAAAGTCGTGGTGTTTGGAATGAAAATACCACCAAATTATGGCACTGCCTA
TAGTCAGGCATTTGAAAATAATTATAAGGTAGTGAGTCAAACATATCAGGTT

AAGTTGTTGCCATTTTTTCTTGATGGTGTGGCTGGACACAAAAGTCTAATGCA
AAATGACCAGATCCATCCAAATGCCAAAGCCCAGTCAATCTTGCTAAATAAC
GCATACCCATATATTAAAGGCGCTTTATAA

F.2.2 Cocos nucifera thioesterase FatB3 (S. cerevisiae codon-optimized, N-term signal peptide removed)

ATGTTGCCAGATTGGTCTATGTTGTTGGCTGCTATTAGAACCATTTTCTCCGCT
GCTGAGAAGCAATGGACTTTGCTCGATTCTAAGAAGCGAGGTGCTGATGCTG
TTGCTGATGCTTCTGGTGTTGGTAAGATGGTTAAGAATGGCTTGGTCTACAGA
CAGAACTTCTCCATTAGATCCTACGAAATTGGTGTTGATAAGAGAGCTTCCGT
TGAGGCTTTGATGAATCATTTCCAAGAAACTTCTTTGAATCATTGTAAGTGTA
TTGGTTTGATGCATGGTGGTTTCGGTTGTACTCCAGAAATGACTAGAAGAAAT
TTGATTTGGGTTGTTGCTAAGATGTTGGTTCATGTTGAAAGATACCCCTGGTG
GGGTGATGTTGTTCAAATTAATACTTGGATTTCTTCTTCTGGTAAGAATGGTA
TGGGTAGAGATTGGCATGTTCATGATTGTCAAACTGGTTTGCCAATTATGAGA
GGTACTTCTGTTTGGGTTATGATGGATAAGCATACTAGAAGATTGTCTAAGTT
GCCAGAAGAAGTTAGAGCTGAAATTACTCCATTCTTCTCTGAAAGAGATGCT
GTTTTGGATGATAATGGTAGAAAGTTGCCAAAGTTCGATGACGATTCTGCTGC
TCATGTTAGAAGAGGTTTGACTCCAAGATGGCATGATTTCGATGTTAATCAAC
ATGTTAATAATGTTAAGTACGTTGGTTGGATTTTGGAATCTGTTCCAGTTTGG
ATGTTGGATGGTTACGAGGTTGCTACTATGTCTTTGGAGTACAGAAGAGAGT
GTAGAATGGATTCTGTTGTTCAATCTTTGACTGCTGTTTCTTCTGATCATGCTG
ATGGTTCTCCAATTGTTTGTCAACATTTGTTGAGATTGGAAGATGGTACTGAA

ATTGTTAGAGGTCAAACTGAATGGAGACCAAAGCAACAAGCTAGAGATTTGG
GTAATATGGGTTTGCATCCAACTGAATCTAAATAA

F.2.3 Cuphea palustris thioesterase FatB1 (S. cerevisiae codon-optimized, N-term signal peptide removed)

ATGAGGCCAAACATGTTGATGGATTCCTTCGGCTTGGAAAGAGTCGTCCAAG
ATGGTTTGGTCTTCAGACAATCCTTCTCCATTAGATCCTATGAAATTTGTGCT
GATAGAACTGCTTCCATTGAAACTGTCATGAACCATGTCCAAGAAACTTCCTT
GAACCAATGTAAGTCCATTGGTTTGTTGGATGATGGTTTCGGTAGATCCCCAG
AAATGTGTAAGAGAGATTTGATTTGGGTCGTCACTAGAATGAAGATTATGGT
CAACAGATACCCAACTTGGGGTGATACTATTGAAGTCTCCACTTGGTTGTCTC
AATCTGGTAAGATTGGTATGGGTAGAGATTGGTTGATTTCTGATTGTAACACT
GGTGAAATTTTGGTCAGAGCTACTTCCGTCTACGCTATGATGAACCAGAAGA
CGAGAAGATTCTCCAAGTTGCCACATGAAGTCAGACAAGAATTTGCTCCACA
TTTCTTGGATTCCCCACCAGCTATTGAAGATAACGATGGTAAGTTGCAAAAGT
TCGATGTCAAGACTGGTGATTCCATTAGAAAGGGTTTGACTCCAGGTTGGTAC
GATTTGGATGTCAACCAACATGTCTCTAACGTCAAGTACATTGGTTGGATTTT
GGAATCTATGCCAACTGAAGTCTTGGAAACTCAAGAATTGTGTTCTTTGACTT
TGGAATACAGAAGAGAATGTGGTAGAGATTCTGTCTTGGAATCCGTCACTTC
TATGGACCCATCTAAGGTCGGTGATAGATTCCAATACAGACATTTGTTGAGAT
TGGAAGATGGTGCTGATATTATGAAGGGTAGAACTGAATGGAGACCAAAGA
ACGCTGGTACTAACGGTGCTATTTCTACTGGTAAGACTTAA

F.2.4 Umbellularia californica thioesterase FatB2 (E. coli codon-optimized, N-term signal peptide removed):

ATGACTCTAGAGTGGAAACCGAAACCAAAACTGCCTCAACTGCTGGATGATC
ACTTCGGTCTGCACGGTCTGGTGTTTCGTCGTACTTTCGCAATTCGTTCTTATG
AAGTGGGTCCAGATCGTTCTACCTCCATCCTGGCCGTCATGAACCACATGCA
GGAAGCCACCCTGAATCACGCGAAATCTGTTGGTATCCTGGGTGATGGTTTC
GGCACTACTCTGGAAATGTCTAAACGTGACCTGATGTGGGTAGTGCGTCGCA
CCCACGTAGCAGTAGAGCGCTACCCTACTTGGGGTGACACTGTGGAAGTCGA
GTGTTGGATTGGCGCGTCCGGTAACAATGGTATGCGTCGCGATTTTCTGGTCC
GTGACTGTAAAACGGGCGAAATCCTGACGCGTTGCACCTCCCTGAGCGTTCT
GATGAACACCCGCACTCGTCGCCTGTCTACCATCCCGGACGAAGTGCGCGGT
GAGATCGGTCCTGCTTTCATCGATAACGTGGCAGTTAAAGACGACGAAATCA
AGAAACTGCAAAAACTGAACGACTCCACCGCGGACTACATCCAGGGCGGTCT
GACTCCGCGCTGGAACGACCTGGATGTTAATCAGCATGTGAACAACCTGAAA
TACGTTGCTTGGGTCTTCGAGACTGTGCCGGACAGCATTTTCGAAAGCCATCA
CATTTCCTCTTTTACTCTGGAGTACCGTCGCGAATGTACTCGCGACTCCGTTCT
GCGCAGCCTGACCACCGTAAGCGGCGGTTCTAGCGAGGCAGGTCTGGTCTGC
GACCATCTGCTGCAACTGGAAGGCGGCTCCGAAGTCCTGCGTGCGCGTACGG
AGTGGCGTCCAAAGCTGACGGATTCTTTCCGCGGCATCTCCGTAATTCCGGCG
GAACCTCGTGTTTAA